DAS
Semantics

Word meaning and close reading

Francis Bond
Department of Asian Studies
Palacký University
https://fcbond.github.io/
bond@ieee.org

Lecture 2

Overview

- Word Meaning
  - Definitions
  - Translations/paraphrases
  - Semantic Relations
  - Components
  - Derivational Relations
  - Word Embeddings
- Wordnet
- Close Reading and Word Sense Disambiguation

The meanings of words

Words carry different meanings: *leave*

10070 *Nothing was left save a few acres of ground , and the two-hundred-year-old house , which is itself crushed under a heavy mortgage .*

10079 *The money which my mother had left was enough for all our wants , and there seemed to be no obstacle to our happiness . ”*

10085 *He had no friends at all save the wandering gipsies , and he would give these vagabonds leave to encamp upon the few acres of bramble- covered land which represent the family estate , and would accept in return the hospitality of their tents , wandering away with them sometimes for weeks on end .*

10107 *She left her room , therefore , and came into mine , where she sat for some time , chatting about her approaching wedding .*

10108 *At eleven o’clock she rose to leave me , but she paused at the door and looked back.*

10439 *” The rest you will leave in our hands . ”*

10449 *And now , Miss Stoner , we must leave you for if Dr. Roylott returned and saw us our journey would be in vain .*

10526 *Then he turned down the lamp , and we were left in darkness .*

How many different meanings?

How can we represent the differences?

- Definitions
- Translations/paraphrases
- Semantic Relations
- Components
- Derivational Relations
- Word Embeddings

Semantic Representations of Words

- Divide meaning into
  - **reference**: the relation to the world/mental space
  - **sense**: the rest of the meaning
    - **denotation** the part that distinguishes the meaning from other meanings
    - **connotation** cultural or emotional associations
- Introduce CONCEPTS
  - How can we represent concepts?
  - How do we learn them?
    - Typically children start off by **underextending** or **overextending** concepts
- Example: *That dog*
  - reference — the animal over there
  - sense — canine quadruped domesticated by man
  - connotation — faithful, friendly (or dirty)

Definitional Semantics

- Standard lexicographic approach to lexical semantics:

  semantics = *the study of language meaning*
  tailor = *a person whose occupation is making and altering garments*

- Definitions are conventionally made up of;
  - genus: what class the lexical item belongs to
  - differentiae: what attributes distinguish it from other members of that class

- Often hard to understand if you don't already know the meaning!

Definitional Semantics: pros and cons

- Pros:
  - familiarity (we are taught to use dictionaries)
- Cons:
  - subjectivity in sense granularity (splitters vs. lumpers) and definition specificity
  - circularity in definitions
  - consistency, reproducibility, …
  - often focus on diachronic (historical) rather than synchronic (current) semantics

Entries for *leave*

02015598-v (72) V1, V2 ***get out, go out, leave, exit*** "move out of or depart from"

02356230-v (8) V3 ***leave, entrust*** "put into the care or protection of someone"

02009433-v (149) V1 ***leave, go away, go forth*** "go away from a place"

02229055-v (7) V3 ***leave, will, bequeath*** "leave or give by will after one's death"

02729414-v (56) V2 ***leave*** "act or be so as to become in a specified state"

02730135-v (5) V2 ***leave*** "have left or have as a remainder"

06690114-n (1) ***leave*** "permission to do something"

Not to be confused with *left hand* and *the leaves fell*, ...

Paraphrases and translation

- Saying the same thing in different words
  - Same language = **paraphrase**
  - Different language = **translation**
- We showed some paraphrases in the entries given above
- If you speak another language, then you can use that to disambiguate may things.
  - ***leave, entrust*** = 預ける *azukeru*
  - ***get out, go out, leave, exit*** = 去る *saru*
  - ***leave, will, bequeath*** = 遺す *nokosu*
- Can you explain the ambiguity in *The money which my mother had left was enough for all our wants*?

Lexical Semantics

- Lexical semantics is concerned with the identification and representation of the semantics of lexical items

- If we are to identify the semantics of lexical items, we have to be prepared for the eventuality of a given word having multiple interpretations
  - **Polysemy**: having multiple meanings
  - **Monosemy**: having only one meaning

- **Homonyms** are words with two unrelated meanings:
  - **homographs**: same spelling
    bow vs *bow*; *keep* vs *keep*
  - **homophones**: same pronunciation
    right vs *write*; *keep* vs *keep*

Distinguishing Polysemes

- **Antagonism**: can the word be used in a sentence with multiple competing interpretations that are incompatible?

  Kim can't bear children

  - Cannot have children
  - Doesn't like children

- **Zeugma**: can the word be used in a context where multiple competing interpretations are simultaneously evoked?

  Kim and her visa expired

  - died
  - ran out

  Hitmen were quite expensive, so she decided to take out a loan and her husband.

- **Paraphrase/Translation**: Is there more than one (clearly different) way to paraphrase/translate the word.

Necessary and Sufficient Conditions

- Can we define words in terms of **conditions**?
  - ***zebra***
    - quadruped
    - animal (redundant)
    - black and white striped
    - herbivore
- These are **intrinsic**, **generic** properties]
  - An albino zebra with three legs is still a zebra
- Can we use words even if we don't know their properties?
  - ***Kway Teow***
- We seem to be ok with fairly vague definitions
  - What is a ***dog-cart***?
  - What is a ***swamp adder***?

Words/Concepts are related in many ways

We can also look at words (or more properly senses) in terms of their relations to other words.

- **Hyponymy/Hypernymy**
- **Synonymy**
- **Antonymy** (Opposites)
- **Meronymy**
  - **Member-Collection**
  - **Portion-Mass**
  - **Element-Substance**
- **Domain** (lexical field)

Hypernymy and Hyponymy

- **Hyponymy**: X is a hyponym of Y iff $f(X)$ entails $f(Y)$ but $f(Y)$ does not entail $f(X)$ (for all or most f):

  Kim has a pet dog $\models$ *Kim has a pet animal*
  Kim has a pet animal $\not\models$ *Kim has a pet dog*

  N.B. complications with universal quantifiers and negation:

  Kim likes all animals $\models$ *Kim likes all dogs*
  Kim likes all dogs $\not\models$ *Kim likes all animals*

- **Hypernymy**: Y is a hypernym of X iff X is a hyponym of Y

- Can a word have multiple hypernyms?

  (1) ***tank***$_1$ $\subset$ ***military_vehicle***$_1$; $\subset$ ***tracked_vehicle***$_1$; $\subset$ ***armored_vehicle***$_1$; ? $\subset$ ***weapon***$_1$

What is entailment

Entailment ($\models$): A sentence p entails a sentence q when the truth of the first (p) guarantees the truth of the second (q), and the falsity of the second (q) guarantees the falsity of the first (p).

Properties of hypernymy/hyponymy

- Asymmetric; applies at the sense level
- applies only to lexical items of the same word class
- Transitive: ***dog***$_1$ $\subset$ ***mammal***$_1$ $\subset$ ***animal***$_1$
- Not all nodes are lexicalized; can be multiple

neutral (Hyper)	male	female	child
sheep	***ram***	***ewe***	***lamb***
cow	***bull***	***cow***	***calf***
goose	***gander***	***goose***	***gosling***
horse	***stallion***	***mare***	***foal:colt/filly***
dog	***dog***	***bitch***	***puppy***
snake	***snake***	***snake***	***snake***

- Can you do this for ***pig***, ***cat*** or ***chicken***?
- Can you give an example of this in another language?

Language Change

- The meanings of words change over time
- ***guitar*** — "a stringed instrument usually having six strings": originally these all used the body to make sound
- We then get ***electric guitar*** — "a guitar with a built-in pickup or pickups which convert string vibrations into electrical signals for amplification"
- To refer to non-electric guitars we get a new coining ***acoustic guitar*** – "a guitar that does not require electrical amplification": which used to just be guitar
  - ***guitar*** is now a hypernym of them both and can refer to either

- Sometimes this practice becomes politically charged, although linguistically it is unremarkable
  - ***woman*** “an adult female person”
  - ***trans woman*** “a person who identifies as a woman but was assigned male at birth”
  - ***cis woman*** “a person who identifies as a woman and was assigned female at birth”
- Can you give other examples of this in English or other languages?

Synonymy

- **Propositional synonymy**: X is a propositional synonym of Y if
  - (i) X and Y are syntactically identical,
  - (ii) substitution of Y for X in a declarative sentence doesn't change its truth conditions

  e.g., ***violin*** and ***fiddle***

- Why propositional synonymy is over-restrictive:
  - syntactic identity (cf. ***eat*** and ***devour***)
  - collocations (cf. ***cemetery*** and ***graveyard***)
  - gradability (cf. ***sofa***/***settee*** vs. ***boundary/frontier***)

Near Synonymy

- Synonyms are substitutable in **some/most** rather than **all** contexts

- Synonymy via semantics: synonyms share "common traits" or attributional overlap, walking the fine line between "necessary resemblances" and "permissible differences":

  grain vs. ***granule***; ***green*** vs. ***purple***; ***alsation*** vs. ***spaniel***

- Permissible differentiation via **clarification**:

  Here is a grain, or granule, of the substance.
  ** The cover is green, {or, that is to say} purple.*

  and **contrast**:

  Here is a grain or, more exactly, granule
  ** He likes alsations, or more exactly, spaniels*

Properties of synonymy

- Symmetric
- traditionally applies only to lexical items of the same word class
  but pairs like ***can*** vs ***be able to*** suggest otherwise
- applied at the sense level?
- $\approx$ converse of polysemy

Antonymy (opposites)

- **Simple antonyms**: the negative of one implies the positive of the other.

  (2) *dead/alive*
  (3) *pass/fail*

- **Gradable Antonyms**: points along a scale

  (4) *boiling/hot/warm/tepid/cool/cold/freezing*
  (5) *like this class/fascinating/interesting/dull/boring*

- **Reverses**: reverse the direction of a motion

  (6) *ascend/descend*
  (7) *up/down; right/left*

- **Converses**: the same act from different points of view

  (8) *above/below; right/left*

  (9) *employer/employee*

  (Slightly non-standard usage by Saeed)

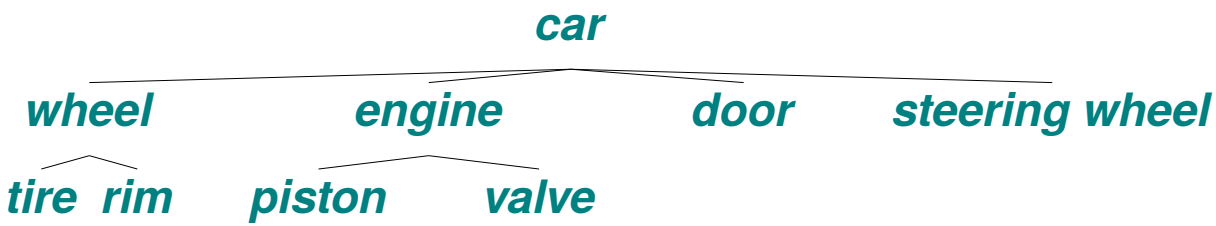
- **Taxonomic Sisters**: children of the same (grand)parent

  (10) *Monday/Tuesday/…/Sunday*
  in WordNet: ***day of the week*** $\supset$ ***weekday***, ***weekend***

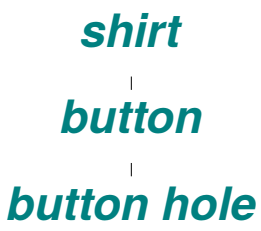
  (11) *LMS/English/Chinese/…*
  Context dependent

Meronymy

- **Meronomy** refers to the part-whole relation
  - **meronym** is the part
  - **holonym** is the whole



- It is not always transitive



But we don't normally say that a ***button hole*** is part of a ***shirt***.

Member-Collection

- The relation between a collection and one of the units that makes it up

(12) *tree–forest*
(13) *sheep–flock*
(14) *fish–school*
(15) *book–library*
(16) *member–band*
(17) *musician–orchestra*
(18) *student–class*

Portion-Mass

- The relation between a mass noun and a typical unit of measurement

  (19) *drop–liquid*
  (20) *grain–sand/salt/truth*
  (21) *sheet/ream–paper*
  (22) *lump–coal (or just about anything)*
  (23) *strand–hair*
  (24) *rasher–bacon*

- Similar to classifiers in many ways, e.g. in Malay

  (25) *ekor* “tail”–*animal*
  (26) *orang* “human”–*person*

Domain (lexical field)

The domain in which a word is typically used with this meaning.

(27) ***driver***$_1$ — the operator of a motor vehicle

(28) ***driver***$_2$ — someone who drives animals that pull a vehicle

(29) ***driver***$_3$ — a golfer who hits the golf ball with a driver [GOLF]

(30) ***driver***$_4$ — ($\simeq$ device driver) a program that determines how a computer will communicate with a peripheral device [COMPUTER SCIENCE]

(31) ***driver***$_5$ — ($\simeq$ number one wood) a golf club (a wood) with a near vertical face that is used for hitting long shots from the tee [GOLF]

Some GOLF terms: ***approach***$_9$, ***approach shot***$_1$, ***golf course***$_1$, ***links course***$_1$, ***wedge***$_5$, ***tee***$_1$, ***scratch***$_9$, ***putt***$_1$, ***slice***$_1$, ***hook***$_1$

And More

- There are many, many more lexical relations advocated by various theories including:
  - Troponymy/hypernymy (cf. *walk* vs. *lollop*) “way of doing something”
  - Entailment (cf. *snore* vs. *sleep*) “if you do one thing, you must be doing the other”
  - Operator (cf. *question* vs. *ask*) “the thing you do by doing something”
  - Magnifier (cf. *wound* vs. *badly*) “intensifier, diminisher”
  - Usage (cf. *strong-willed* vs. *pig-headed* “stubborn”)
    pig-headed is **pejorative**

Componential Analysis

Break word meaning into its components

- For example:

woman	[FEMALE]	[ADULT]	[HUMAN]	
spinster	[FEMALE]	[ADULT]	[HUMAN]	[UNMARRIED]
bachelor	[MALE]	[ADULT]	[HUMAN]	[UNMARRIED]
wife	[FEMALE]	[ADULT]	[HUMAN]	[MARRIED]
girl	[FEMALE]	[CHILD]	[HUMAN]	
boy	[MALE]	[CHILD]	[HUMAN]	

semantic components/**primitives** shown as [COMPONENT]

- components allow a compact description
- interact with morphology/syntax
- form part of our cognitive architecture

Defining Relations using Components

- **hyponymy**

  A lexical item P is a hyponym of Q if all the components of Q are also in P.

woman	[FEMALE]	[ADULT]	[HUMAN]	
spinster	[FEMALE]	[ADULT]	[HUMAN]	[UNMARRIED]
wife	[FEMALE]	[ADULT]	[HUMAN]	[MARRIED]

  spinster $\subset$ ***woman***; ***wife*** $\subset$ ***woman***

- **incompatibility**

  A lexical item P is incompatible with Q if they share some components but differ in one or more **contrasting** components

  spinster $\not\approx$ ***wife***

Binary Features

- We can make things more economical (fewer components):

woman	[+FEMALE]	[+ADULT]	[+HUMAN]	
spinster	[+FEMALE]	[+ADULT]	[+HUMAN]	[–MARRIED]
bachelor	[–FEMALE]	[+ADULT]	[+HUMAN]	[–MARRIED]
wife	[+FEMALE]	[+ADULT]	[+HUMAN]	[+MARRIED]
girl	[+FEMALE]	[-ADULT]	[+HUMAN]	

  - Which should be $+$? [+FEMALE] or [–MALE]
  - Presumably also [–ELECTRIC], [–CONICAL], …
    Only show **relevant** features
  - **antonyms** differ in only one binary component

Redundancy Rules

- We can add relations between components:

[+HUMAN]	→	[+ANIMATE]	
[+ADULT]	→	[+ANIMATE]	
[+ANIMATE]	→	[+CONCRETE]	
[+MARRIED]	→	[+ADULT]	
[+MARRIED]	→	[+HUMAN]	...

- Which allows us to write:

woman	[+FEMALE]	[+ADULT]	[+HUMAN]	
spinster	[+FEMALE]	[+ADULT]	[+HUMAN]	[–MARRIED]
bachelor	[–FEMALE]	[+ADULT]	[+HUMAN]	[–MARRIED]
wife	[+FEMALE]			[+MARRIED]

Can we say [–MARRIED] → [+HUMAN]?

More Complex Breakdowns

- We can add relations between components:

[+FATHER]	→	[+MALE] [+PARENT]
[+FATHER](x,y)	→	[+MALE](x) [+PARENT](x,y)
[+SON](x,y)	→	[+MALE](x) [+PARENT](y,x)
[+BROTHER](x,y)	→	[+MALE](x) [+PARENT](z,x) [+PARENT](z,y)
[+GRANDFATHER](x,y)	→	[+MALE](x) [+PARENT](x,z) [+PARENT](z,y)

- Assume [+PARENT](x,y) means "x is the parent of y"

- There are various ways you can formalize such relationships
  - Many parts of language can be formalized in such a way
  - We did this for demonstratives
    this, that, these, those, what, here, there, where

These are great for many sub-systems of language, but it is hard to make components for everything, …

Derivational Relations

- There are also relations between meanings with different parts of speech
  Here we say one word is derived from the other

- Normally these have some morphological component
  - ***beauty*** (n) — ***beautiful*** (adj) — ***beautifully*** (adv) — ***beautify*** (verb)
  - ***decision*** (n) —***decide*** (v) —***decisive*** (adj) —***decisively*** (adv)
  - ***drive*** (v) —***driver*** (n) —***drivable*** (adj)
  - ***creation*** (n) —***create*** (v) —***creative*** (adj) —***creatively*** (adv)
  - *Sandy cut their hand* — *Sandy has a cut on their hand*

Agentive Nouns

- An **agentive noun** is a word that is typically derived from another word denoting an action, and that identifies an entity that does that action.
  verb + ***-er, -or, -ant***

  (32) ***murderer, commentator, whaler, director, computer***

  (33) ?? ***undertaker, cooker, footballer*** (Saeed also includes these)

- Should ***murderer*** be listed separately from ***murder*** in the dictionary? Why or why not?

- Also **recipient noun**s that show the undergoer: **verb** + ***-ee***: ***employee***, ***trustee***

Agentive Nouns in Other Languages

- Japanese (suffix distinguishes person/machine)
  - 運転する → 運転者 *unten-sha* “driver”
  - 計算する → 計算者 計算機 *keisan-sha/ki* “computer”
  - 研究する → 研究者 研究員 *kenkyuu-sha/in* “researcher”
  - 読む → 読み手 読者 *yomite/dokusha* “reader”
- Malay (prefix can convert any part of speech)
  - *bantu (v)* “help” → *pembantu* “assistant/helper”
  - *potong (v)* “cut” → *pemotong* “cutter (human/machine)”
  - *terbang (v)* “fly” → *penerbang* “pilot (not passenger)”

Word Embeddings

- Represent words as a vector of numbers (instead of a set of components)
- Every word has a unique word embedding (or "vector"), which is just a list of numbers for each word.
- Embeddings start being useful from 50-500 dimensions
  LLMs typically are much larger
- The embedding captures the "meaning" of the word.
- Similar words end up with similar embedding values
- Context based word embeddings give a different vector depending on the context

Word Embeddings

- In the simplest case, each word is a number

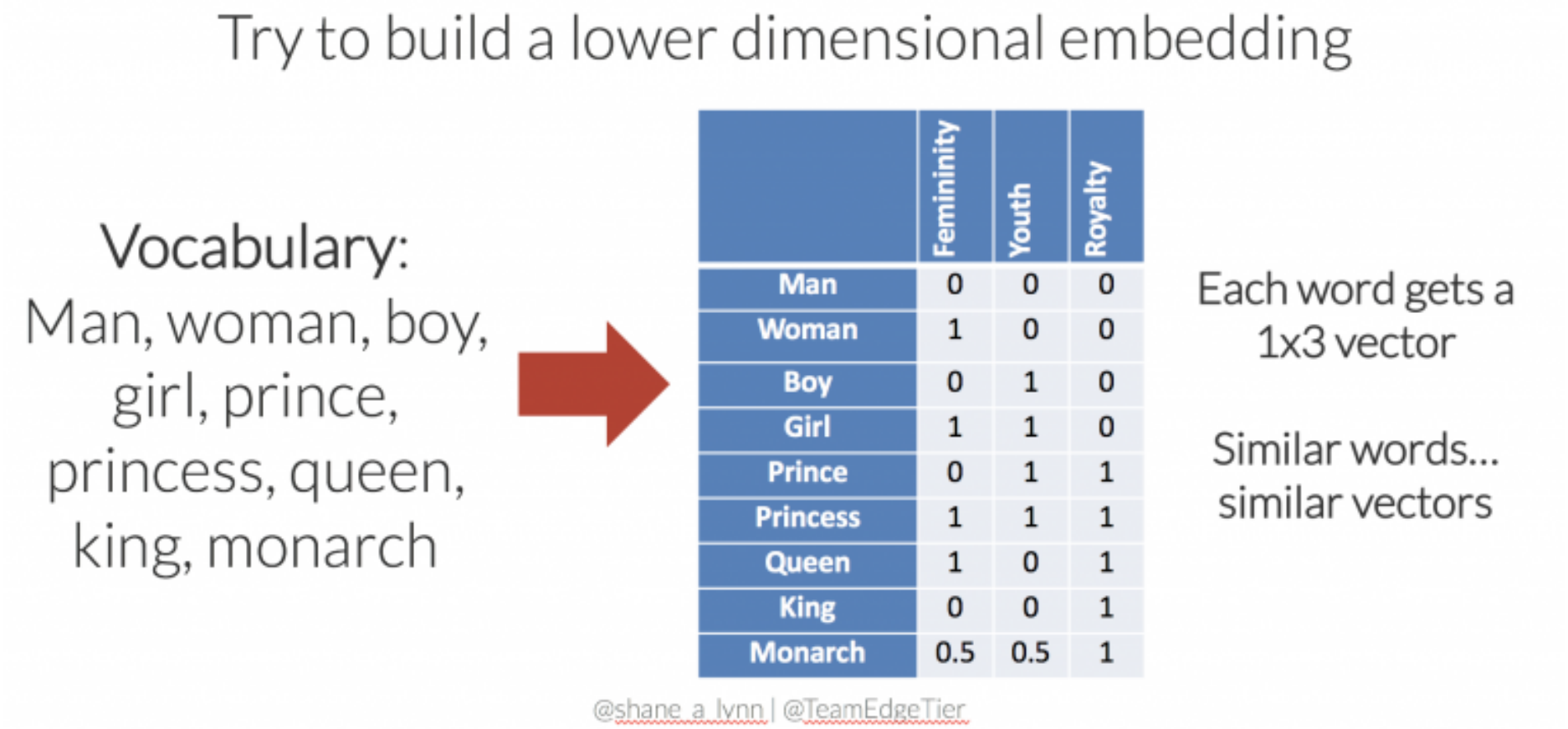
Vocabulary:
Man, woman, boy, girl, prince, princess, queen, king, monarch

	1	2	3	4	5	6	7	8	9
man	1	0	0	0	0	0	0	0	0
woman	0	1	0	0	0	0	0	0	0
boy	0	0	1	0	0	0	0	0	0
girl	0	0	0	1	0	0	0	0	0
prince	0	0	0	0	1	0	0	0	0
princess	0	0	0	0	0	1	0	0	0
queen	0	0	0	0	0	0	1	0	0
king	0	0	0	0	0	0	0	1	0
monarch	0	0	0	0	0	0	0	0	1

Each word gets a 1x9 vector representation

- Two many dimensions
- No shared information
- Mainly zeros

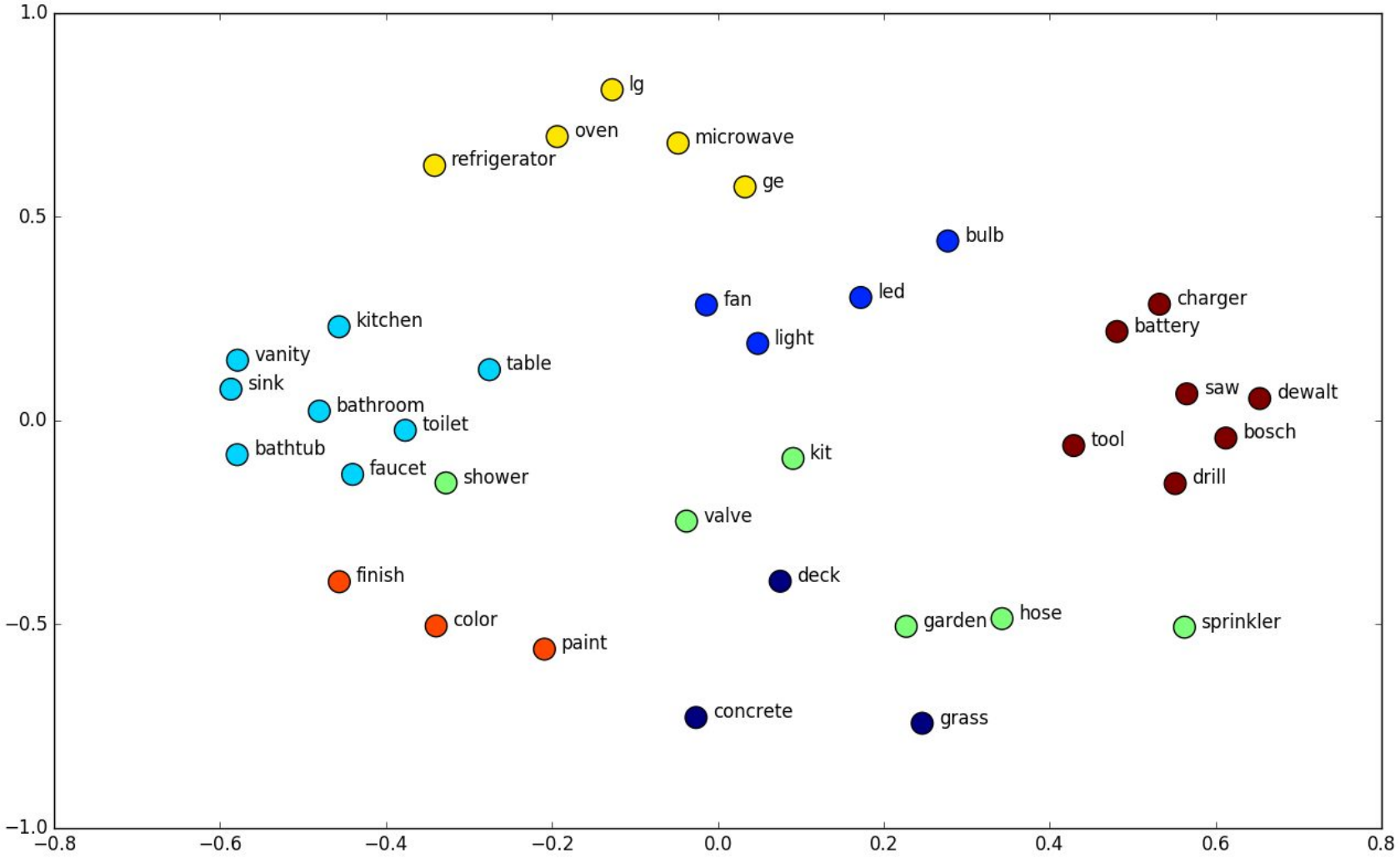
- We want fewer, more meaningful, dimensions (like components!)



	Femininity	Youth	Royalty
Man	0	0	0
Woman	1	0	0
Boy	0	1	0
Girl	1	1	0
Prince	0	1	1
Princess	1	1	1
Queen	1	0	1
King	0	0	1
Monarch	0.5	0.5	1

- How would you add *child*? or *emperor*?

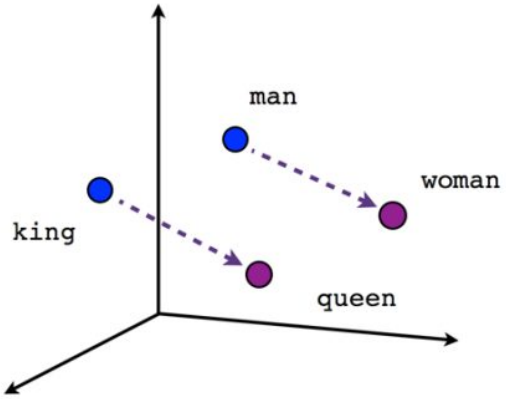
Similar words should be close



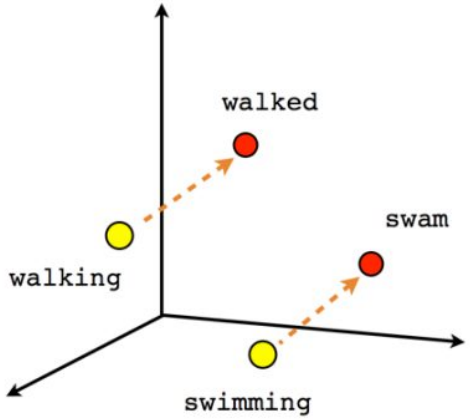
We can learn these from raw text

- In LLMs, models are constructed to predict the context words from a centre word, or the centre word from a set of context words.
- By training on large amounts of text, embeddings that model human intuitions can be built.
- By using multilingual text, we can link languages
- More about this in **DAS/LTI**: Language Technology and the Internet

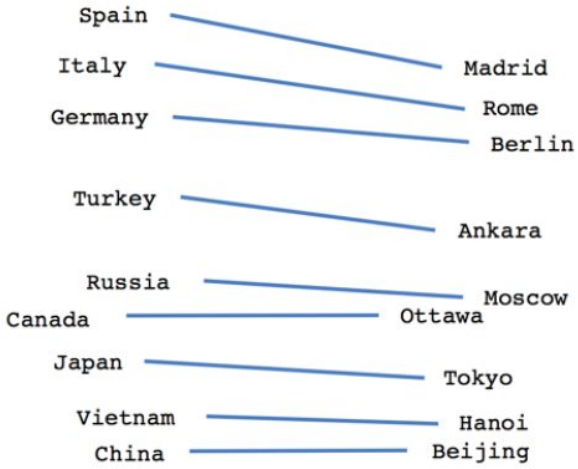
Semantic relations are also learned



Male-Female



Verb tense



Country-Capital

We can do arithmetic on the vectors

$$\vec{king} + \vec{woman} - \vec{man} \approx \vec{queen}$$

$$\vec{Paris} - \vec{France} + \vec{Germany} \approx \vec{Berlin}$$

Corpora contain stereotypes, ML learns them!

- We can test if things are closer to $\vec{he}$ or $\vec{she}$

$$
\begin{array}{lcr}
\vec{nurse}.\vec{she} & = & 0.38 \\
\vec{nurse}.\vec{he} & = & -0.12 \\
\vec{programmer}.\vec{she} & = & 0.07 \\
\vec{programmer}.\vec{he} & = & 0.28
\end{array}
$$

- This is an accurate description of the state of the world described in the corpus

- But may not be what we want to use as a basis for reasoning, …

Wordnet

WordNet

- WordNet is an open-source electronic lexical database of English, developed at Princeton University

  http://wordnet.princeton.edu/

- Made up of four separate semantic nets, for each of nouns, verbs, adjectives and adverbs

- WordNets exist for many languages, my group has worked on:
  - Japanese
  - Bahasa Malay/Indonesian
  - Chinese (Mandarin and Cantonese)
  - The shared open multi-lingual wordnet (150+ languages)
    https://omwn.org/
  - Kristang
  - Myanmar
  - Czech

Wordnet Structure

- Lexical items are categorised into ∼115K (and counting) glossed **synsets** (= synonym sets)

```
1. enrichment -- (act of making fuller or more
   meaningful or rewarding)
2. enrichment -- (a gift that significantly increases
   the recipient's wealth)
```

- Lexical relations at either the synset level or sense (= combination of lexical item and synset) level

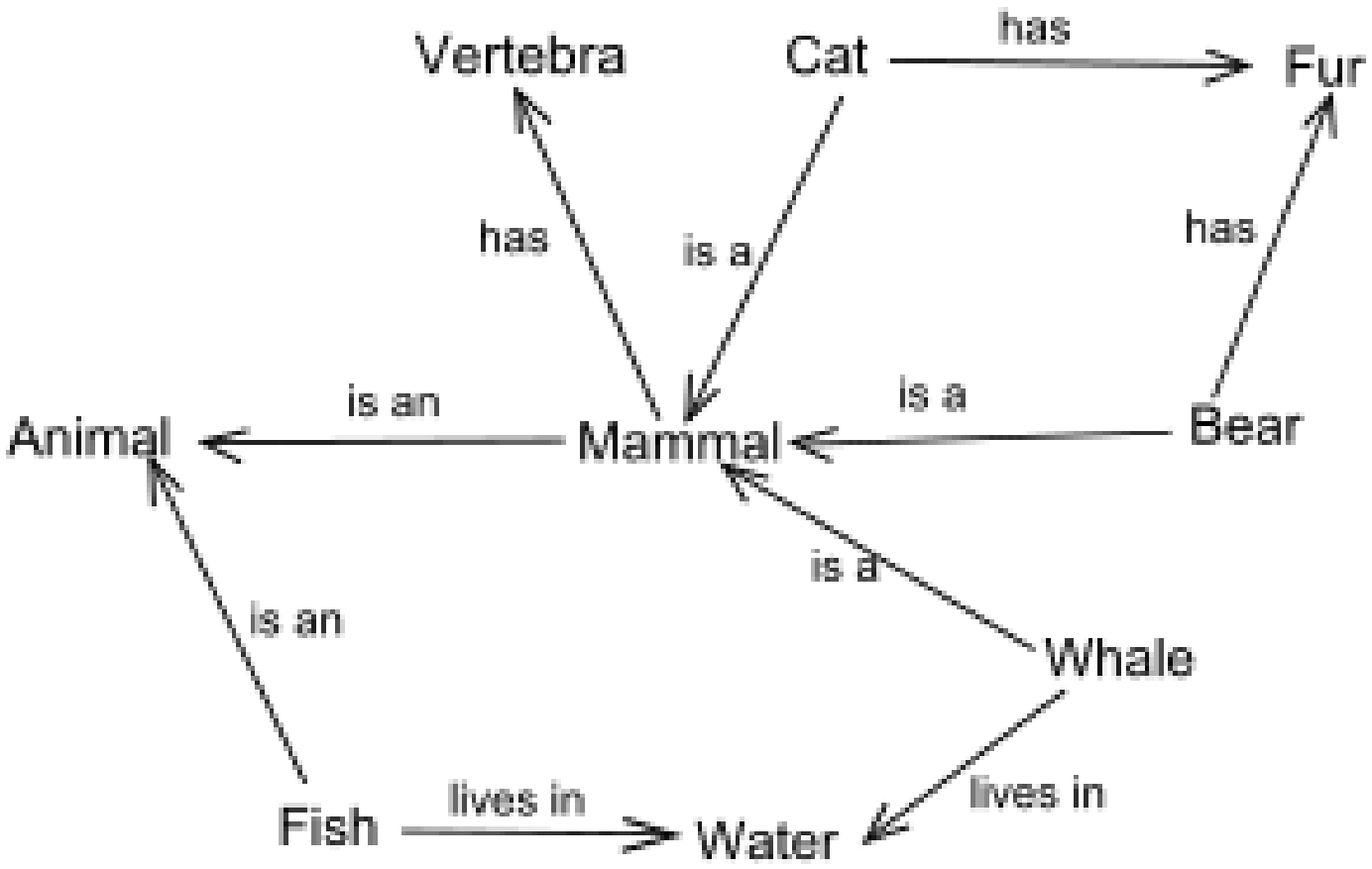
- Strongly lexicalist (orginally):
  - synsets only where words exist
  - but many multiword expressions ($\approx 50\%$)

Psycholinguistic Foundations of WordNet

- Strong foundation on hypo/hypernymy (lexical inheritance) based on
  - response times to sentences such as:
    a canary {can sing/fly,has skin}
    a bird {can sing/fly,has skin}
    an animal {can sing/fly,has skin}
  - analysis of anaphora:
    I gave Kim a novel but the {book,?product,...} bored her
    Kim got a new car. It has shiny {wheels,?wheel nuts,...}
  - selectional restrictions

- Is now often used to calculate **semantic similarity**
  - The shorter the path between two synsets the more similar they are
  - Or the shorter the path to the nearest shared hypernym, …

Word Meaning as a Graph



- You need a very big graph to capture all meanings

Wordnet in this course

- We will use wordnet to test our skills in determining word meaning
  - tag a short text from this year's story or stories
  - discuss differences with other annotators
- As well as a source of examples and inspiration

Word Sense Disambiguation for Close Reading

Close Reading

- Reading (and often re-reading) a text to uncover multiple aspects of meaning that lead you to understand a text better
- Looking at what the text actually says, as well as the inferences you make from reading it
- After a close reading you should be able to support your conclusions with specific examples from the text
- You can consider many aspects of the text, such as
  - The Title
  - Word Choice
  - The Tone and Style
  - Discerning Patterns
  - Point of View and Characterization
  - Symbolism

The idea comes from the study of poetry and is part of the school of *New Criticism* (e.g., Brooks and Warre, 1938, 1943)

Word Choice and Diction

- What word(s) stand out? Why? (typically vivid words, unusual choices, or a contrast to what a reader expects)
- How do particular words get us to look at characters or events in a particular way? Do they evoke an emotion?
- Did the author use nonstandard language or words in another language? Why? What is the effect?
- Are there any words that could have more than one meaning? Why might the author have played with language in this way?
- Do some words have extra connotations?

Word Sense Disambiguation

- Knowing what individual words mean is the first step towards understanding
- We will try to identify the **sense** of words
  - We use Wordnet (Fellbaum, 1998) as the sense inventory
    because it contains semantic relations as well as definitions
    and it is accessible: there are good interfaces to it
  - For every word we chose the most appropriate sense in wordnet
    or write a comment if we think there isn't one
  - Once we have identified a sense, it is then easy to look at synonyms and other closely related words
  - We use the Czech wordnet (Pala and Smrž, 2004) for Czech
  - Both wordnets have been extended as part of the the Natural Text Understanding — Multilingual Corpus (**ntu-mc**)

WSD: Example

Sherlock Holmes had been leaning back in his chair with his eyes closed and his head sunk in a cushion, but he half opened his lids now and glanced across at his visitor.

1. ***hat, chapeau, lid*** “headdress that protects the head from bad weather; has shaped crown and usually a brim”
2. ***lid, eyelid, palpebra*** “either of two folds of skin that can be moved to cover or open the eye”
3. ***lid*** “a movable top or cover (hinged or separate) for closing the opening at the top of a box, chest, jar, pan, etc.”

WSD: A challenging task!

This is difficult for many reasons

- Meaning boundaries are not clear: the sense distinctions impose a structure on something that is actually fuzzy
- Dictionaries are imperfect
  - senses may be missing
  - senses may be too fine-grained
- Processing a text by computer is difficult
  - The computer may have misinterpreted
    - the part-of-speech
      Does that go “Female deer which go” “Is it the case that is goes?”
      The speckled band “the band that is speckled” “the band that someone speckled”
    - Or the sentence boundaries
    - Or the words boundaries
- People use language idiosyncratically
  - extending meanings metaphorically
  - sometimes so strangely that we might even say wrongly

WSD: not impossible

- Typically people agree around 72.5% of the time (Snyder and Palmer, 2004).
  - Verbs are hardest (67.8%), then nouns (74.9%) and adjectives (78.5%)
  - Disagreements tend to cluster around a relatively small group of difficult words.
  - For example ***national***
    - In six out of seven instances one annotator chose “limited to or in the interests of a particular nation” and the other annotator chose “concerned with or applicable to or belonging to an entire nation or country”
    - They are hard to distinguish!

Projects 1 and 2

1 Identify and annotate word meaning for your own passage of one of the stories using wordnet as the sense inventory: 3 or 4 people do the same set of sentences.

- For every word that needs to be tagged, either
  - Chose a sense in wordnet
  - Identify it as a named entity
  - Identify a problem in the corpus or wordnet
    and leave a comment saying what the it should be
- Do this on your own — the goal is to think about the words' meanings

2 Compare and contrast your annotations with other annotators; re-annotate based on your discussion and leave comments for at least five words.

- We expect you to disagree 30-40% of the time (more often than experts)
- Sometimes you will have made a careless mistake
- Sometimes you will have interpreted wordnet differently
- Sometimes you will have interpreted the passage differently
- Discussing meaning deepens your understanding of it

Conclusions

- We learned a little about word meaning, wordnet, close reading and your assignment
- This is covered in more detail in Saeed (2015, Chapter 3)

Acknowledgments and References

- Definitions from WordNet: http://wordnet.princeton.edu/
- Images from
  - the Open Clip Art Library: http://openclipart.org/
  - Steven Bird, Ewan Klein, and Edward Loper (2009) *Natural Language Processing with Python*, O'Reilly Media
    www.nltk.org/book
- Video: Dead parrot sketch by Monty Python

*

References

Tolga Bolukbasi, Kai-Wei Chang, James Zou, Venkatesh Saligrama, and Adam Kalai. 2016. Man is to computer programmer as woman is to homemaker? debiasing word embeddings. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, NIPS'16, page 4356–4364. Curran Associates Inc., Red Hook, NY, USA.

Cleanth Brooks and Robert Penn Warre. 1938. *Understanding Poetry*. Appleton-Century-Crofts.

Cleanth Brooks and Robert Penn Warre. 1943. *Understanding Fiction*. Appleton-Century-Crofts.

Christine Fellbaum, editor. 1998. *WordNet: An Electronic Lexical Database*. MIT Press.

George Miller. 1998. Foreword. In Fellbaum (1998), pages xv–xxii.

Karel Pala and Pavel Smrž. 2004. Building Czech wordnet. *Romanian Journal of Information Science*, 7:79–88.

John I. Saeed. 2015. *Semantics*. Wiley-Blackwell, fourth edition.

Benjamin Snyder and Martha Palmer. 2004. The English all-words task. In *Proceedings of Senseval-3*, pages 41–44. ACL, Barcelona.