LECTURE 7: LEXICAL SEMANTICS

Mehmet Can Yavuz, PhD.

Adapted from Julia Hockenmaier, NLP S2023 - course material
https://courses.grainger.illinois.edu/cs447/sp2023/

LET'S LOOK AT WORDS AGAIN….

So far, we've looked at…

… the **structure** of words (**morphology**)

… the **distribution** of words (**language modeling**)

Today, we'll start looking at the **meaning** of words (**lexical semantics**).

We will consider:

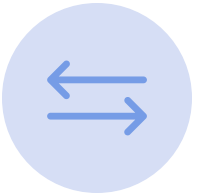
… the **distributional hypothesis** to identify words with similar meanings

… two kinds of **vector representations** of words that are inspired by the distributional hypothesis.

TODAY'S LECTURE

Part 1: Lexical Semantics and the Distributional Hypothesis

Part 2: Distributional similarities (from words to sparse vectors)



Part 3: Word embeddings(from words to dense vectors)

WHAT DO WORDS **MEAN**, AND HOW DO WE **REPRESENT** THAT?

Do we want to represent that…

… “cassoulet” is a French dish?

… “cassoulet” contains meat?

… “cassoulet” is a stew?

WHAT DO WORDS **MEAN,** AND HOW DO WE **REPRESENT** THAT?

Do we want to represent…

… that a “bar” is a place to have a drink?

… that a “bar” is a long rod?

… that to “bar” something means to block it?

DIFFERENT APPROACHES TO LEXICAL SEMANTICS

Roughly speaking, NLP draws on two different types of approaches to capture the meaning of words:

- **The lexicographic tradition** aims to capture the information represented in lexicons, dictionaries, etc.
- **The distributional tradition** aims to capture the meaning of words based on large amounts of raw text

THE LEXICOGRAPHIC TRADITION

Uses resources such as lexicons, thesauri, ontologies etc. that capture explicit knowledge about word meanings.

Assumes words have *discrete* word senses:

bank1 = financial institution; bank2 = river bank, etc.

May capture *explicit relations* between word (senses): "*dog*" is a "*mammal*", "*cars*" have "*wheels*" etc.

THE DISTRIBUTIONAL TRADITION

Uses large corpora of raw text to learn the meaning of words from the contexts in which they occur.

Maps words to (sparse) vectors that capture corpus statistics

Contemporary variant: use neural nets to learn dense vector "embeddings" from very large corpora
(this is a prerequisite for most neural approaches to NLP)

If each word type is mapped to a single vector, this ignores the fact that words have multiple senses or parts-of-speech.

LEXICOGRAPHIC APPROACHES TO WORD MEANING

WHERE WE'RE AT

We have looked at how to represent the meaning of sentences based on the meaning of their words (using predicate logic).

Now we will get back to the question of how to represent the meaning of words (although this won't be in predicate logic)

We will look at lexical resources (WordNet) We will consider two different tasks:

- Computing word similarities
- Word sense disambiguation

DIFFERENT APPROACHES TO LEXICAL SEMANTICS

Lexicographic tradition (today's lecture)

- Use lexicons, thesauri, ontologies
- Assume words have discrete word senses:
  - bank1 = financial institution; bank2 = river bank, etc.
- May capture explicit relations between word (senses): "dog" is a "mammal", etc.

•**Distributional tradition** (earlier lectures)

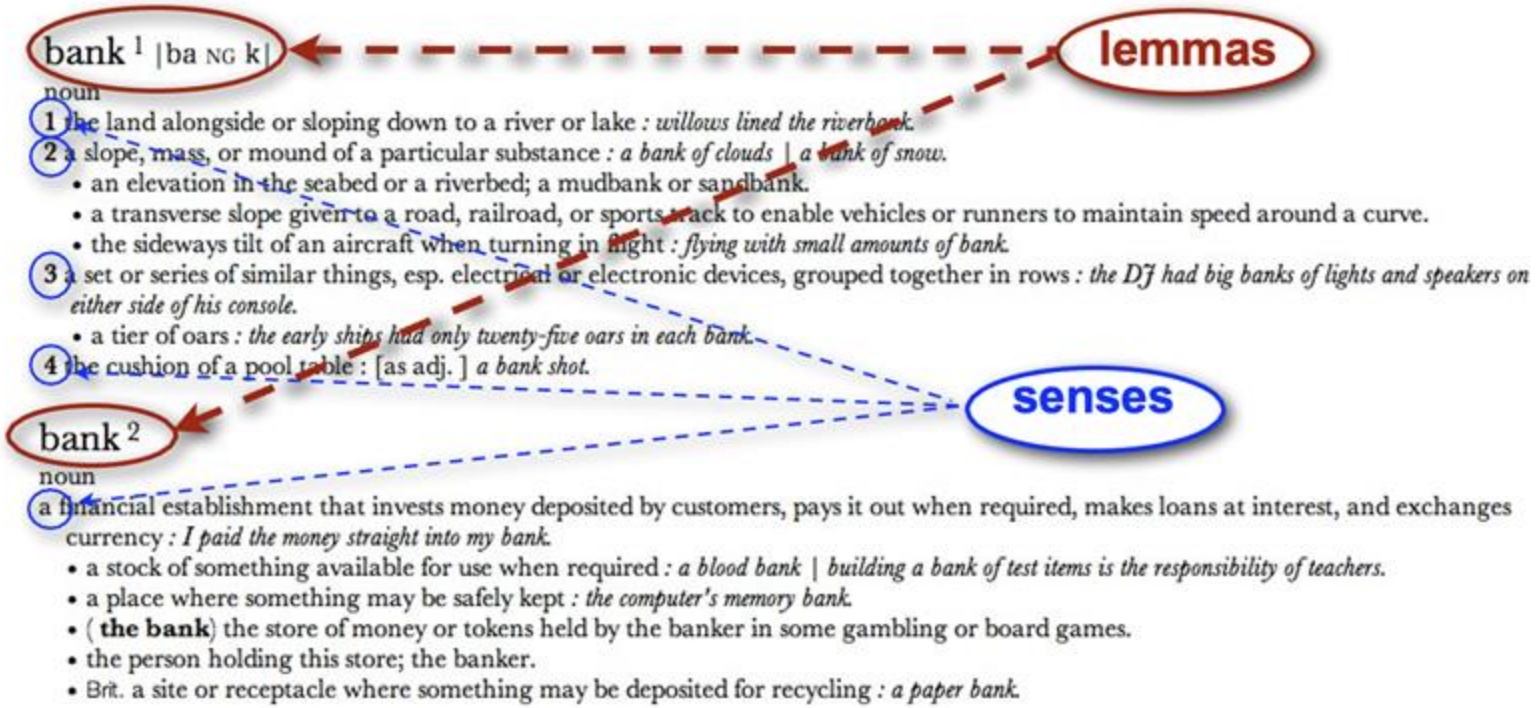
- Map words to (sparse) vectors that capture corpus statistics
- Contemporary variant: use neural nets to learn dense vector "embeddings" from very large corpora
  - (this is a prerequisite for most neural approaches to NLP)
- This line of work often ignores the fact that words have multiple senses or parts-of-speech

WORD SENSES

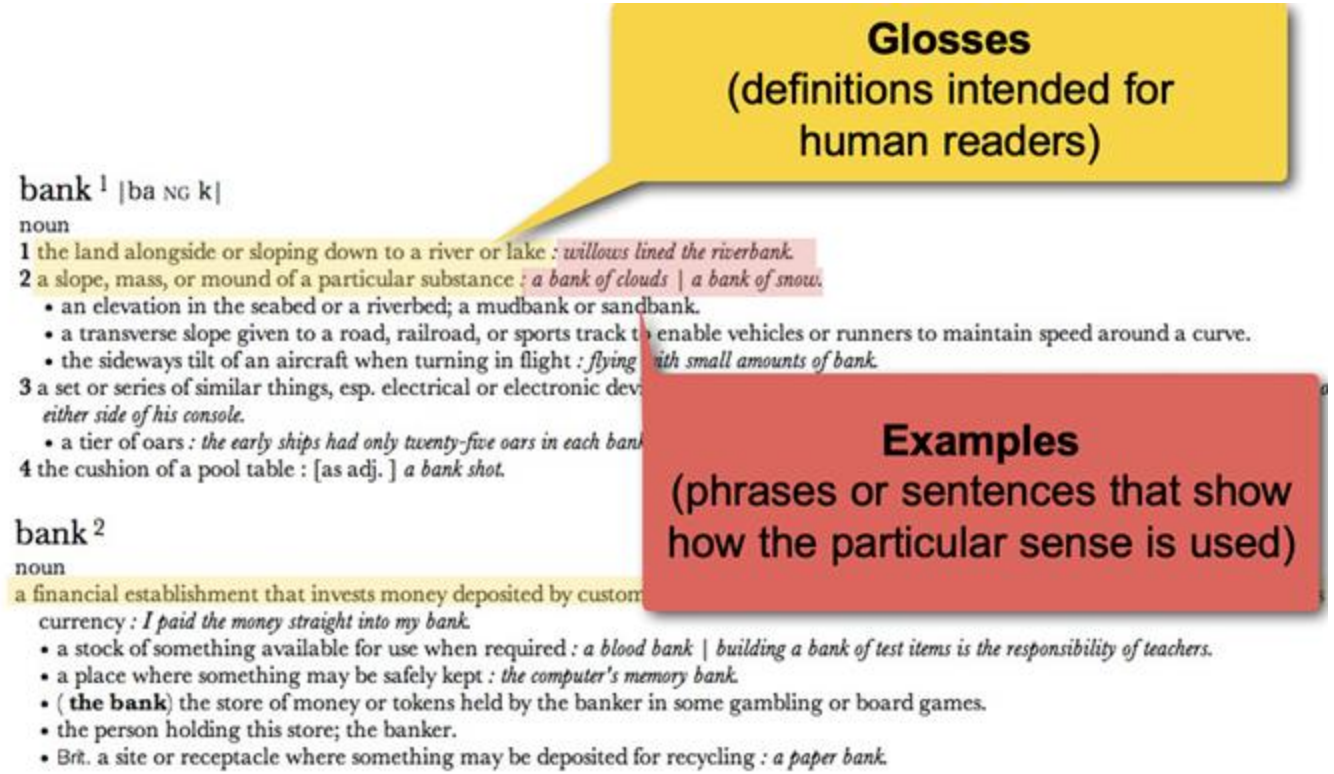
What does *'bank'* mean?

- 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



LEXICON ENTRIES



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)

Lemma (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)

Lexeme: 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

TRYING TO MAKE SENSE OF SENSES

Polysemy:

A lexeme is polysemous if it has different *related senses*

Busey

bank = financial institution or building

Homonyms:

Two lexemes are homonyms if their *senses are unrelated*, but they happen to have the **same spelling and pronunciation**

bank = (financial) bank or (river) bank

RELATIONS BETWEEN SENSES

Symmetric relations:
Synonyms: *couch/sofa*
Two lemmas with the **same** sense

Antonyms: *cold/hot, rise/fall, in/out*

Two lemmas with the **opposite** sense

Hierarchical relations:

Hypernyms and **hyponyms**: *pet/dog*

The hyponym *(dog)* is **more specific** than the hypernym *(pet)*

Holonyms and **meronyms:** *car/wheel*

The meronym *(wheel)* is a **part of** the holonym *(car)*

METONYMY

Some senses of a word may be related in a systematic way, e.g. ...

... organizations and buildings:

I see you in front of the bank on Green Street.

... cars and their drivers:

This Camry looks new. vs. *The Camry honked at me.*

... authors and their works:

Jane Austen wrote Emma. vs *I really like Austen*

... plants and the food derived from them:

Plums have beautiful blossoms. vs *I ate a plum*

WORDNET AND WORDNET-BASED WORD SIMILARITY

WORDNET

•Very large, publicly available **lexical database** of English:
110K nouns, 11K verbs, 22K adjectives, 4.5K adverbs

(WordNets for many other languages exist or are under construction)

•Each word has a POS tag and one or more **word senses**.
Avg. # of senses: 1.23 nouns, 2.16 verbs, 1.41 adj, 1.24 adverbs

•Word senses are grouped into synonym sets ("**synsets**")

81K noun synsets, 13K verb synsets, 19K adj. synsets, 3.5K adverb synsets

•Synsets are connected in a hierarchy/network defined
via **conceptual-semantic relations**
- hypernym/hyponym relation (IS-A)
- holonym/meronym relation (HAS-A)

Also lexical relations (derivational morphology), and lemmatization

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

A WORDNET EXAMPLE
Searching for “bass” returns

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"*

HIERARCHICAL SYNSET RELATIONS: NOUNS (I)

IS-A relations (hyponymy):

Hypernym/hyponym (between concepts)

- *meal* is a hypernym (superordinate) of *breakfast breakfast* is a hyponym (subordinate) of *meal dog* is a hypernym (superordinate) of *poodle poodle* is a hyponym (subordinate) of (IS-A) *dog*

Instance hypernym/hyponym (concepts and instances)

- *composer* is the instance hypernym of (HAS-INSTANCE) *Bach Bach* is an instance hyponym of (IS-INSTANCE-OF) *composer*

WORDNET HYPERNYMS AND HYPONYMS

(n) **bass** (the lowest part of the musical range)
- *direct hypernym* / *inherited hypernym* / *sister term*
  - S: (n) pitch (the property of sound that varies with variation in the frequency of vibration)
    - S: (n) sound property (an attribute of sound)
      - S: (n) property (a basic or essential attribute shared by all members of a class) *"a study c*
        - S: (n) attribute (an abstraction belonging to or characteristic of an entity)
          - S: (n) abstraction, abstract entity (a general concept formed by extracting cc
            - S: (n) entity (that which is perceived or known or inferred to have its

(n) **bass**, bass part (the lowest part in polyphonic music)
- *direct hyponym* / ***full hyponym***
  - S: (n) ground bass (a short melody in the bass that is constantly repeated)
  - S: (n) figured bass, basso continuo, continuo, thorough bass (a bass part written out in full and accomp
- *direct hypernym* / *inherited hypernym* / *sister term*
  - S: (n) part, voice (the melody carried by a particular voice or instrument in polyphonic music) *"he trie*
    - S: (n) tune, melody, air, strain, melodic line, line, melodic phrase (a succession of notes forming
      - S: (n) music (an artistic form of auditory communication incorporating instrumental or vo
        - S: (n) auditory communication (communication that relies on hearing)
          - S: (n) communication (something that is communicated by or to or between
            - S: (n) abstraction, abstract entity (a general concept formed by extracti
              - S: (n) entity (that which is perceived or known or inferred to hav

HIERARCHICAL SYNSET RELATIONS: NOUNS (II)

Part-Whole relations (meronymy):

Member holonym/meronym (groups and members)

- *crew* is a member holonym of (HAS-MEMBER) co-pilot
- *co-pilot* is a member meronym of (IS-MEMBER-OF) *crew*

Part holonym/meronym (wholes and parts)

- *car* is a part holonym of (HAS-PART) *wheel wheel* is a part meronym of (IS-PART-OF) *car*

Substance holonym/meronym (substances and components)

- *bread* is a substance holonym of (HAS-COMPONENT) *flour flour* is a substance meronym of (IS-COMPONENT-OF) *bread*

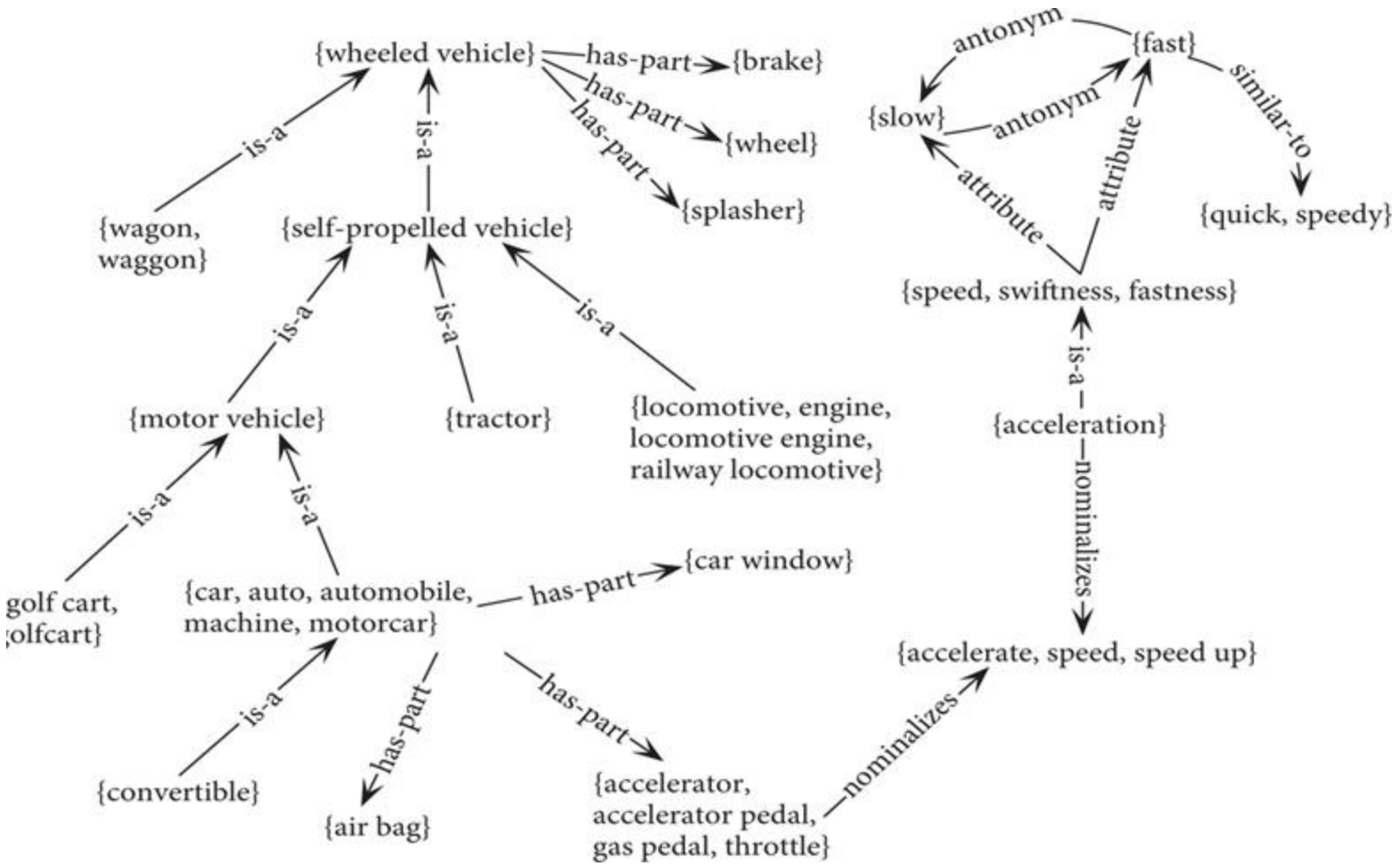
HIERARCHICAL SYNSET RELATIONS: VERBS

Hypernym/troponym (between events):

- *travel/fly, walk/stroll*
- *Flying* is a troponym of *traveling:*
- it denotes **a specific manner** of *traveling*

Entailment (between events):

- *snore/sleep*
- *Snoring* **entails (presupposes)** *sleeping*
- (if somebody is snoring, they have to be sleeping)



(Figure from Jurafsky & Martin, 3rd Edition, and Navigli 2016)

WORDNET RELATIONS AS A GRAPH

WORDNET AS A SEMANTIC NETWORK

The **Hypernym/hyponym** relations (IS-A) and **holonym/meronym** relations (HAS-A) in WordNet capture some important world knowledge, e.g.:

- car IS-A motor-vehicle IS-A... IS-A wheeled-vehicle wheeled-vehicle HAS-A brake
  - ® car IS-A wheeled-vehicle
  - ® car HAS-A brake

We can interpret WordNet as a simple "semantic network" (for semantic networks in AI see e.g. http:// www.jfsowa.com/pubs/semnet.htm)

WORDNET-BASED WORD SIMILARITY

There have been many attempts to exploit resources like WordNet to compute word (sense) similarities.

Classic approaches use the distance (**path length**) between synsets (these paths typically only consider hypernym/hyponym relations), possibly augmented with corpus statistics

More recent (neural) approaches aim to learn (non-Euclidean) embeddings that capture the hierarchical hypernym/hyponym structure of WordNet.

WHAT DO WE MEAN BY "WORD (SENSE) SIMILARITY"?

There are many aspects to "similarity":

— **Similarity as synonymy:**

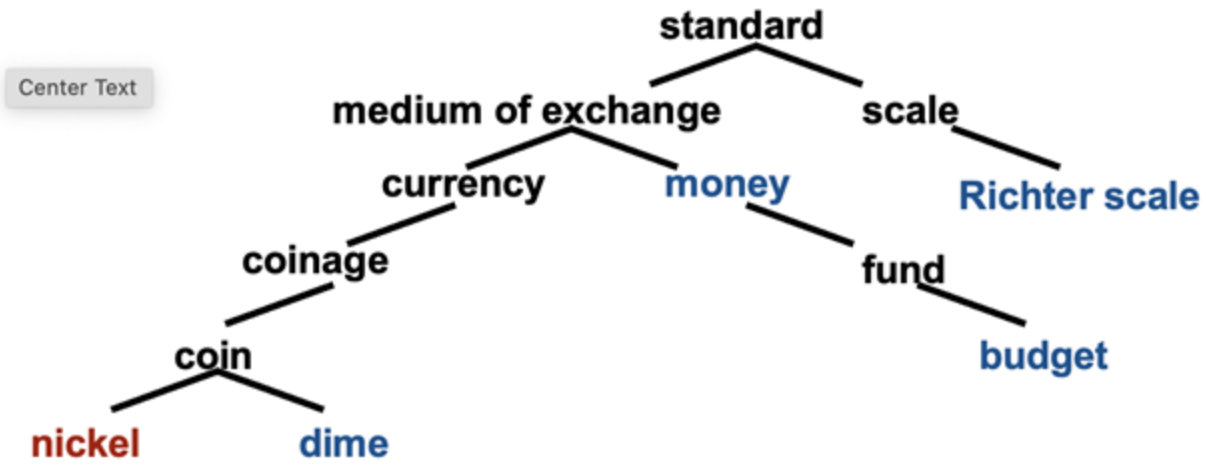
- sim(*couch*, *sofa*)> sim(*poodle*, *dog*) > sim(*poodle*, *pug*), … Do the two words/senses have the same meaning?
- (WordNet: synsets are synonyms (similarity=1), but hypernym/hyponyms (*dog*/*poodle*) are also more similar to each other than unrelated words)

— **Similarity as association:**

- How related are the two words/senses to each other? *coffee* and *cup* are strongly associated, but not synonyms "**Semantic fields**": sets of words that are topically related
- (WordNet: holonyms/meronyms etc. capture some associations)

Earlier metrics of similarity in NLP often conflate both notions, but see e.g. SimLex-999 https://www.aclweb.org/anthology/J15-4004.pdf

WORDNET PATH LENGTHS: EXAMPLES AND PROBLEMS



Path length is just the distance between synsets

pathlen(nickel, dime) = 2 (nickel—coin—dime)

pathlen(nickel, money) = 5 (nickel—…—medium of exchange—money)

pathlen(nickel, budget) = 7 (nickel—…—medium of exchange—…–budget)

But do we really want the following?

pathlen(nickel, coin) < pathlen(nickel, dime)

pathlen(nickel, Richter scale) = pathlen(nickel, budget)

PROBLEMS WITH THESAURUS-BASED SIMILARITY

We need to have a thesaurus! (not available for all languages)

We need to have a thesaurus that contains the words we're interested in.

We need a thesaurus that captures a rich hierarchy of hypernyms and hyponyms.

Most thesaurus-based similarities depend on the specifics of the hierarchy that is implement in the thesaurus.

LEARNING HYPONYM RELATIONS

If we don't have a thesaurus, can we *learn* that Corolla is a kind of car from text?

Certain **phrases and patterns** indicate hyponym relations:

- Hearst(1992) [Hearst patterns]
- **Enumerations:** *cars* ***such as*** *the Corolla, the Civic, and the Vibe,*
- **Appositives:** *the Corolla* ***,*** *a popular car…*

We can also **learn these patterns** if we have some **seed examples of hyponym relations** (e.g. from WordNet):

1. *Take all hyponym/hypernym pairs from WordNet (e.g. car/vehicle)*
2. *Find all sentences that contain both, and identify patterns*
3. *Apply these patterns to new data* to get new hyponym/hypernym pairs