Natural Language Processing

03: Senses and WordNet

Philipp Schaer, Technische Hochschule Köln, Cologne, Germany

Technology
Arts Sciences
TH Köln

Polysemy

- … the movie is so bad, in fact, that it retains that ridiculous tarzan call that was so tirelessly mocked in last summer's comedy "george of the jungle."
- … the performances are bad.

Polysemy

	Sense 1	Sense 2
bad	Of poor quality or little worth.	slang (orig. U.S.). Formidable, good.
sick	Suffering from illness of any kind; ill, unwell, ailing.	slang (now esp. Skateboarding sick and Surfing). Excellent, impressive; risky.
awesome	Inspiring awe; appalling, dreadful, weird.	Expressing enthusiastic approval: great, excellent, highly impressive; fantastic.
wicked	Bad in moral character, disposition, or conduct	Excellent, splendid; remarkable. slang (orig. U.S.).

Oxford English Dictionary

Word senses

I'm going to the bank

- bank$_1$ = "financial institution"

- bank$_2$ = "sloping mound"

- bank$_3$ = "biological repository"

- bank$_4$ = "building where a bank$_1$ does its business"

A word sense is a representation of one **aspect** of a word's meaning.

Word senses – "serve"

Dictionary
2 found

A | A

serve

All   English   **English Thesaurus**   German   German - English   Apple   Wikipedia

serve

verb

1 *they have served their political masters faithfully for the past 40 years*: **WORK FOR**, be in the service of, perform duties for, be employed by, have a job with; obey, be obedient to, carry out the wishes of.

2 *I decided that I wanted to work somewhere where I could serve the community*: **BE OF SERVICE TO**, be of use to, help, give help to, assist, give assistance to, aid, lend a hand to, give a helping hand to, do a good turn to, make a contribution to, do one's bit for, do something for, benefit; minister to, succour.

3 *altogether she had served on the committee for 11 years*: **BE A MEMBER OF**, work on, be on, sit on, have/hold a place on, perform duties on, carry out duties on.

4 *California is limiting the number of terms a politician can serve in office* | *Lewis served his apprenticeship in Scotland*: **CARRY OUT**, perform, do, fulfil, complete, discharge; spend, go through.

5 *serve the soup hot with lots of crusty bread* | *dinner is served at candlelit tables*: **DISH UP/OUT**, give out, distribute, set out, plate up, spoon out, ladle out; present, provide, supply, make available.

6 *Elizabeth walked off to serve another customer*: **ATTEND TO**, give one's attention to, attend to the requirements of, deal with, see to; **ASSIST**, help, look after, take care of.

7 *the landlord's daughter served at table*: **ACT AS WAITER/WAITRESS**, wait, distribute food/refreshments; *North American informal* sling hash, sling plates.

8 *they were just about to serve him with a writ*: **DELIVER TO**, present with, give to, hand over to, cause to accept.

9 *she stabbed the cigarette out in a saucer serving as an ashtray*: **ACT AS**, function as, fulfil the function of, do duty as, do the work of, act as a substitute for.

10 *official forms are obtainable that, with minor adaptation, will serve in all but a few cases*: **SUFFICE**, be adequate, be good enough, be all right, fit/fill the bill, do, answer; be useful, serve a purpose, meet requirements, suit.

Word senses

- They rarely serve red meat

- He served as U.S. ambassador to Norway

- He might have served his time.

Zeugma

- **Conjunction** ("yoke") of **antagonistic** readings

- one test for whether word senses are distinct.

  - Which flights serve breakfast?

  - Does Midwest Express serve Philadelphia?

  - Does Midwest Express serve breakfast and Philadelphia?

Relationship between senses

- Synonymy

- Antonymy

- Hypernymy

- Meronymy

- And many others… like Metonymy

Synonym

- Two senses of different words are synonyms of each other if their meaning is nearly identical

- Two words are never exactly the same in their meaning, distribution of use, dialect or other contexts in which they're licensed.

- Synonyms can be exchanged for each other without changing the truth conditions of a sentence.

- Synonymy holds between word senses, not words

couch	sofa
filbert	hazelnut
car	automobile
fair	impartial
fair	pale

Antonymy

- Two senses of different words are antonyms of each other if their meaning is nearly opposite

- All aspects of meaning are nearly identical between antonyms, except one (very much like synonyms in this respect)

long	short	both describe length
big	little	both describe size
fast	slow	both describe speed
cold	hot	both describe temperature
dark	light	both describe luminescence

Hyponymy

hypo = "under"
(e.g., hypothermia)

- Sense A is a hyponym of sense B if A is a subclass of B
- Formally, entailment: for entity x, $A(x) \Rightarrow B(x)$
- Hyponymy is generally transitive
  - dog -> mammal -> animal

hyponym/subordinate	hypernym/superordinate
car	vehicle
mango	fruit
chair	furniture
dog	mammal
mammal	animal

Meronymy

- Part-whole relations. A meronym is a part of a holonym.

meronym	holonym
leg	chair
wheel	car
car	automobile

WordNet

- Lexical database for nouns, verbs and adjectives/abverbs.

- Each word sense is arranged in a synset (category of near-synonyms) and each synset is related to others in terms of their sense relations.

- General information: https://wordnet.princeton.edu

- Online interface: http://wordnetweb.princeton.edu/perl/webwn

Relations

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 their instances	$composer^1 \rightarrow Bach^1$
Part Meronym	Has-Part	From wholes to parts	$table^2 \rightarrow leg^3$
Part Holonym	Part-Of	From parts to wholes	$course^7 \rightarrow meal^1$
Antonym		Semantic opposition between lemmas	$leader^1 \Longleftrightarrow follower^1$
Derivation		Lemmas w/same morphological root	$destruction^1 \Longleftrightarrow destroy^1$

Figure 18.3 Some of the noun relations in WordNet.

Image source: Jurafsky and Martin (2021), p. 361

Synsets

synset	gloss
mark, grade, score	a number or letter indicating quality
scratch, scrape, scar, mark	an indication of damage
bell ringer, bull's eye, mark, home run	something that exactly succeeds in achieving its goal
hump, fool, gull, mark, patsy, fall, guy, sucker, soft touch, mug	a person who is gullible and easy to take advantage of
mark, stigma, brand, stain	a symbol of disgrace or infamy

Hypernyms of bass

```
bass³, 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
bass⁷ (member with the lowest range of a family of instruments)
=> musical instrument, instrument
   => device
      => instrumentality, instrumentation
         => artifact, artefact
            => whole, unit
               => object, physical object
                  => physical entity
                     => entity
```

Figure 18.5 Hyponymy chains for two separate senses of the lemma *bass*. Note that the chains are completely distinct, only converging at the very abstract level *whole, unit*.
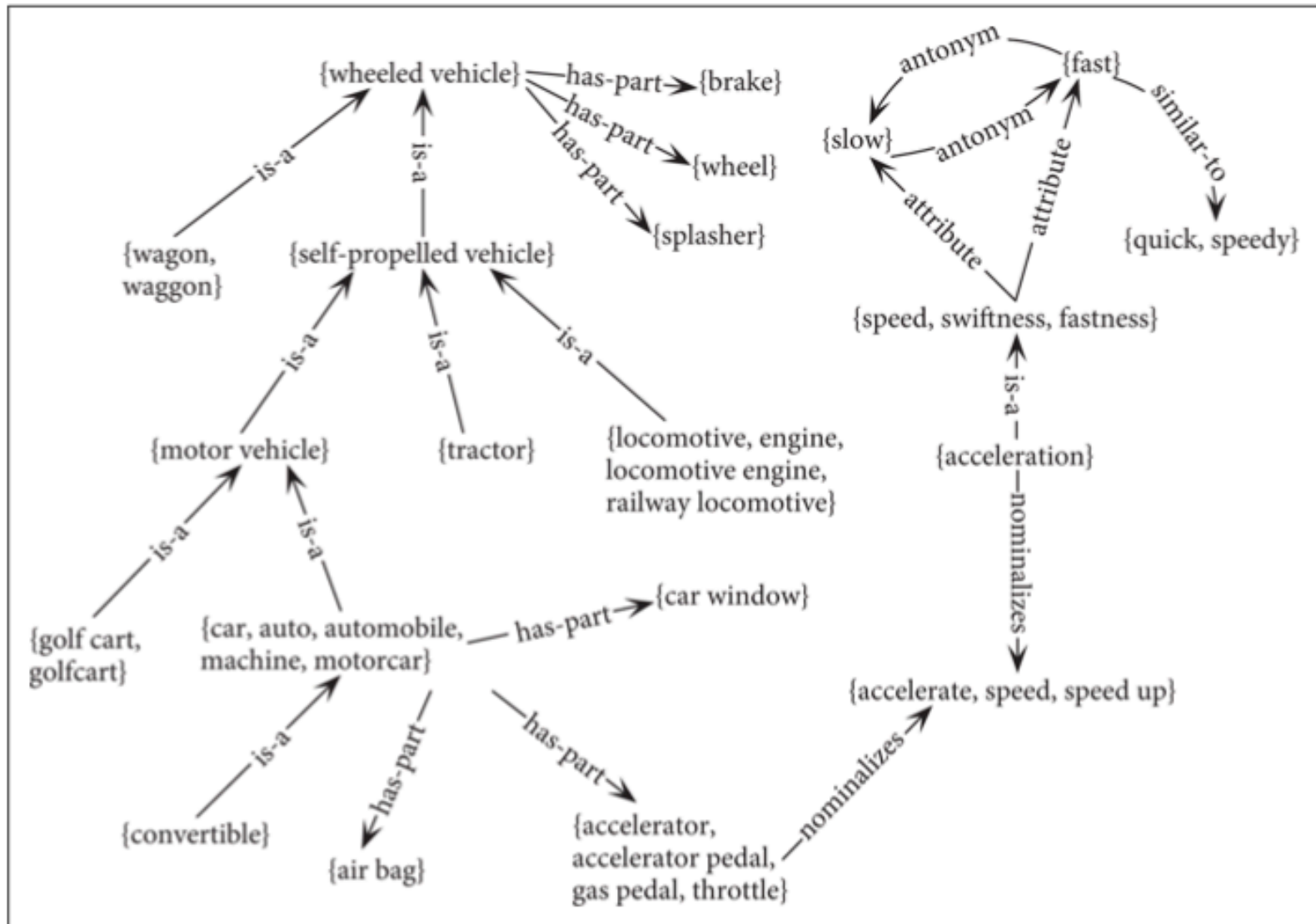
Image source: Jurafsky and Martin (2021), p. 362

Figure 18.6 WordNet viewed as a graph. Figure from Navigli (2016).

Image source: Jurafsky and Martin (2021), p. 362

WordNet

- WordNet encodes human-judged measures of similarity.

- The WordNet 3.0 release has 117,798 nouns, 11,529 verbs, 22,479 adjectives, and 4,481 adverbs. The average noun has 1.23 senses, and the average verb has 2.16 senses.

- By indexing word senses, we can build annotated resources on top of it for word sense disambiguation.

Word sense disambiguation

"All-word" WSD

- "Only$_{only1}$ a relative$_{relative1}$ handful$_{handful1}$ of such$_{such0}$ reports$_{report3}$ was received$_{receive2}$"

- For all content words in a sentence, resolve each token to its sense in an fixed sense inventory (e.g., WordNet).

- Simple but strong heuristics:
  - Choose the most frequent sense (equivalent, in WordNet, to take the first sense)
  - One sense per discourse

All-word WSD example
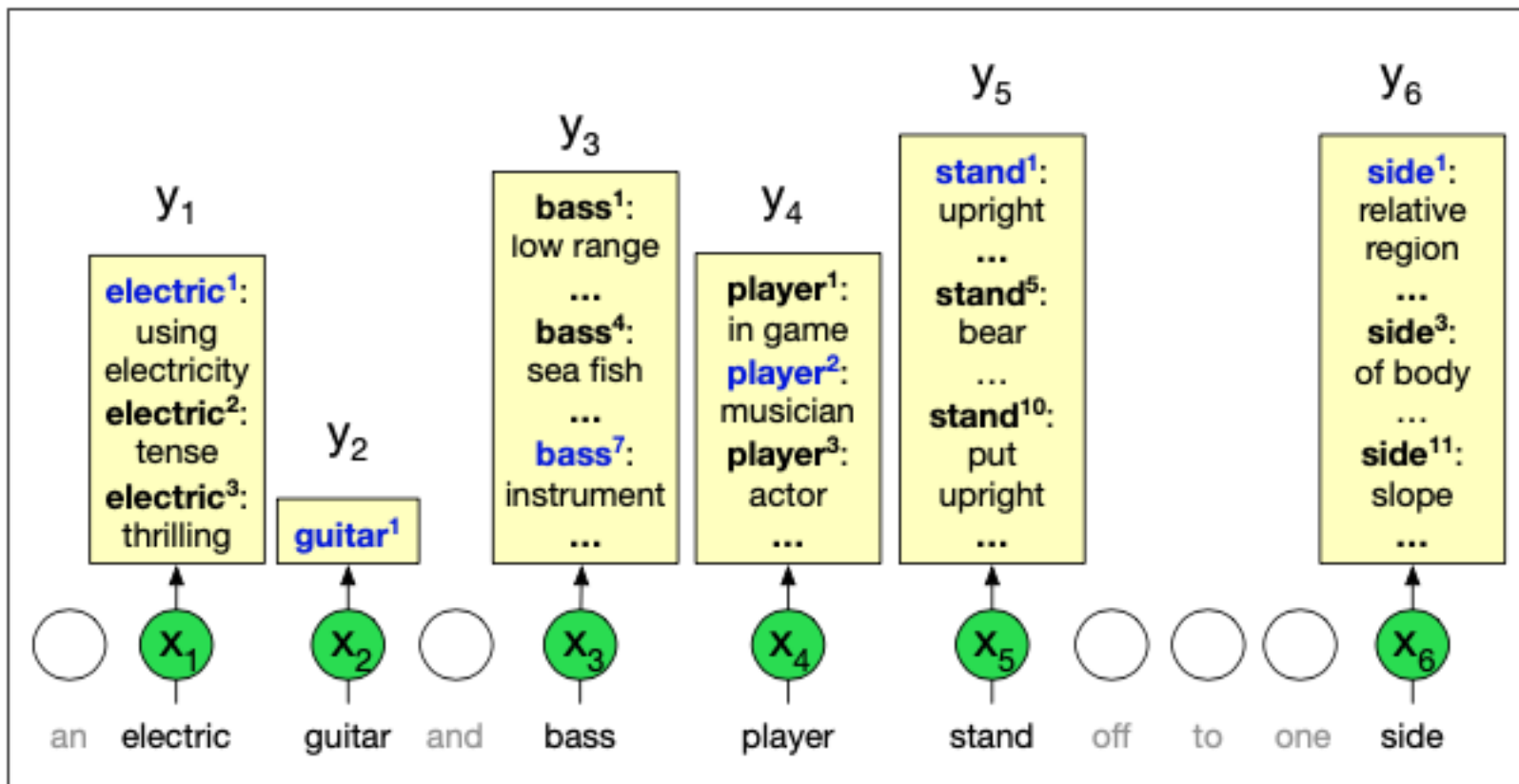
Dictionary WSD

- Predict the sense a given token that has the sense's dictionary highest overlap between the token's context and sense's dictionary gloss.

bank$_1$	Gloss	a financial institution that accepts deposits and channels the money into lending activities
	Examples	"he cashed a check at the bank", "that bank holds the mortage on my home"
bank$_2$	Gloss	sloping land (especially the slope beside a body of water)
	Examples	"the pulled the canie up on the bank", "he sat on the bank of the river and watched the currents"

- "The boat washed up on the river bank."

Lesk Algorithm

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

   best-sense ← most frequent sense for *word*

   max-overlap ← 0

   context ← set of words in *sentence*

   for each *sense* **in** senses of *word* **do**

     signature ← set of words in gloss and examples of *sense*

     overlap←COMPUTEOVERLAP(*signature,context*)

     if *overlap* > *max-overlap* **then**

       max-overlap ← *overlap*

       best-sense←*sense*

   end

   return(*best-sense*)

Lesk Algorithm

- Knowledge-based algorithms, rely solely on WordNet or other such resources and don't require labeled data

- Simplified Lesk algorithm.
  - The COMPUTEOVERLAP function returns the number of words in common between two sets, ignoring function words or other words on a stop list.
  - The original Lesk algorithm defines the *context* in a more complex way.

- There are many obvious extensions to Simplified Lesk, such as weighing the overlapping words by IDF to downweight frequent words like function words

Other Applications of WordNet

- ✓ word-sense disambiguation

- information retrieval

- automatic text classification

- automatic text summarization

- machine translation

- and even automatic crossword puzzle generation

A common use of WordNet is to check similarity between words

- measure distance between synsets in WordNet's graph
- NTLK: `synset1.path_similarity(synset2)`
- https://www.nltk.org/howto/wordnet.html

Mental homework

- Using your favourite dictionary, simulate the original Lesk word overlap disambiguation algorithm on the phrase *Time flies like an arrow*. Assume that the words are to be disambiguated one at a time, from left to right, and that the results from earlier decisions are used later in the process.

- How can we use WordNet in the ESUPOL scenario?