

Lexical and Vector Semantics

CSE538 - Spring 2023
Natural Language Processing

Topics

- Lexical Ambiguity (why word sense disambiguation)
- Word Vectors
- Topic Modeling

Objectives

- Define common semantic tasks in NLP.
- Understand linguistic information necessary for semantic processing.
- Learn a couple approaches to semantic tasks.
- Motivate deep learning models necessary to capture language semantics.

Terminology: lemma and wordform

- A **lemma** or **citation form**
 - Same stem, part of speech, rough semantics
- A **wordform**
 - The inflected word as it appears in text

Wordform	Lemma
banks	bank
sung	sing
duermes	dormir

(Jurafsky & Martin, SLP, 2019)

Lemmas have senses

- One lemma “bank” can have many meanings:

Sense 1: • ...a **bank**₁ can hold the investments in a custodial account...

Sense 2: • “...as agriculture burgeons on the east **bank**₂ the river will shrink even more”

- **Sense (or word sense)**

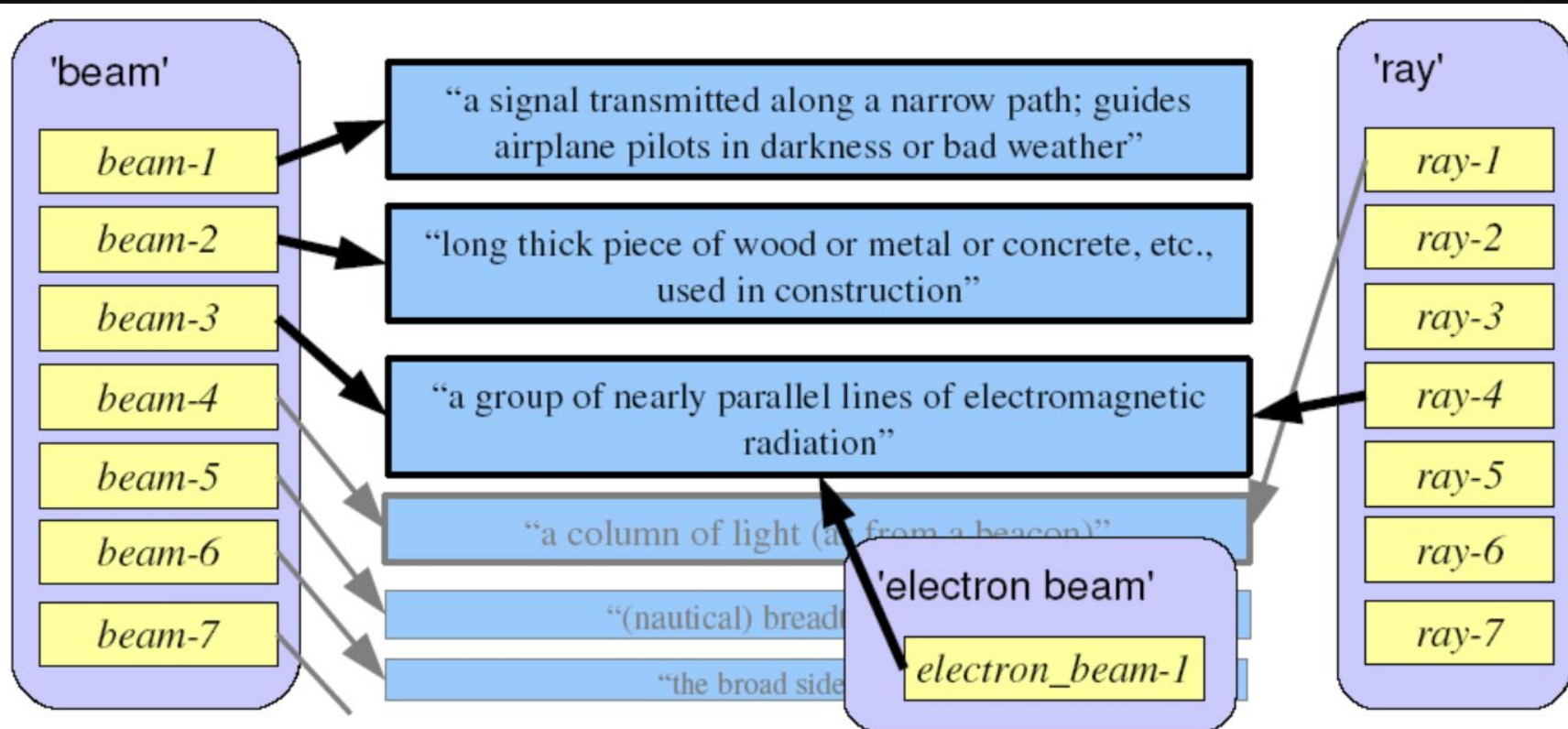
- A discrete representation

of an aspect of a word’s meaning.

- The lemma **bank** here has two senses

(Jurafsky & Martin, SLP, 2019)

Lemmas have senses



(Schwartz, 2011)

Homonymy

Homonyms: words that share a form but have unrelated, distinct meanings:

- bank₁: financial institution, bank₂: sloping land
- bat₁: club for hitting a ball, bat₂: nocturnal flying mammal

1. Homographs (bank/bank, bat/bat)

2. Homophones:

1. Write and right
2. Piece and peace

(Jurafsky & Martin, SLP, 2019)

Homonymy causes problems for NLP applications

- Information retrieval
 - “bat care”
- Machine Translation
 - bat: murciélago (animal) or bate (for baseball)
- Text-to-Speech
 - bass (stringed instrument) vs. bass (fish)

(Jurafsky & Martin, SLP, 2019)

Word Sense Disambiguation

He put the **port** on the ship.

He walked along the **port** of the steamer.

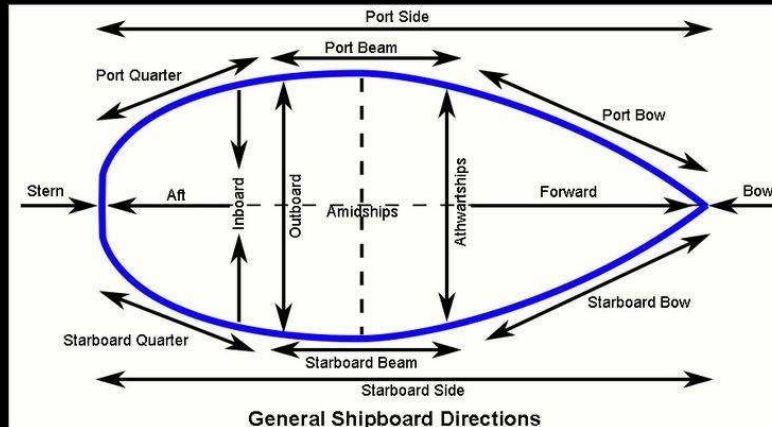
He walked along the **port** next to the steamer.

Word Sense Disambiguation

He put the **port** on the ship.

He walked along the **port** of the steamer.

He walked along the **port** next to the steamer.

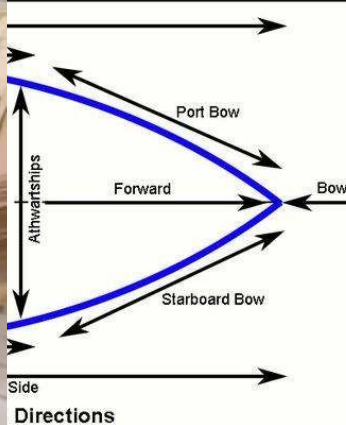


Word Sense Disambiguation

He put the **port** on the ship.

He walked along the **port** of the steamer.

He walked along the **port** next to the steamer.



Word Sense Disambiguation

He put the **port** on the ship.

He walked along the **port** of the steamer.

He walked along the **port** next to the steamer.

port.n.1 (a place (seaport or airport) where people and merchandise can enter or leave a country)

port.n.2 port wine (sweet dark-red dessert wine originally from Portugal)

Word Sense Disambiguation

He put the **port** on the ship.

He walked along the **port** of the steamer.

He walked along the **port** next to the steamer.

port.n.1 (a place (seaport or airport) where people and merchandise can enter or leave a country)

port.n.2 port wine (sweet dark-red dessert wine originally from Portugal)

port.n.3, embrasure, porthole (an opening (in a wall or ship or armored vehicle) for firing through)

Word Sense Disambiguation

He put the **port** on the ship.

He walked along the **port** of the steamer.

He walked along the **port** next to the steamer.

port.n.1 (a place (seaport or airport) where people and merchandise can enter or leave a country)

port.n.2 port wine (sweet dark-red dessert wine originally from Portugal)

port.n.3, embrasure, porthole (an opening (in a wall or ship or armored vehicle) for firing through)

larboard, **port.n.4** (the left side of a ship or aircraft to someone who is aboard and facing the bow or nose)

Word Sense Disambiguation

He put the **port** on the ship.

He walked along the **port** of the steamer.

He walked along the **port** next to the steamer.

port.n.1 (a place (seaport or airport) where people and merchandise can enter or leave a country)

port.n.2 port wine (sweet dark-red dessert wine originally from Portugal)

port.n.3, embrasure, porthole (an opening (in a wall or ship or armored vehicle) for firing through)

larboard, **port.n.4** (the left side of a ship or aircraft to someone who is aboard and facing the bow or nose)

interface, **port.n.5** ((computer science) computer circuit consisting of the hardware and associated circuitry that links one device with another (especially a computer and a hard disk drive or other peripherals))

Word Sense Disambiguation

He put the **port** on the ship.

He walked along the **port** of the steamer.

He walked along the **port** next to the steamer.

As a verb...

1. **port** (put or turn on the left side, of a ship) *"port the helm"*
2. **port** (bring to port) *"the captain ported the ship at night"*
3. **port** (land at or reach a port) *"The ship finally ported"*
4. **port** (turn or go to the port or left side, of a ship) *"The big ship was slowly porting"*
5. **port** (carry, bear, convey, or bring) *"The small canoe could be ported easily"*
6. **port** (carry or hold with both hands diagonally across the body, especially of weapons) *"port a rifle"*
7. **port** (drink port) *"We were porting all in the club after dinner"*
8. **port** (modify (software) for use on a different machine or platform)

port.n.1 (a place (seaport or airport) where people and merchandise can enter or leave a country)

port.n.2 port wine (sweet dark-red dessert wine originally from Portugal)

port.n.3, embrasure, porthole (an opening (in a wall or ship or armored vehicle) for firing through)

larboard, **port.n.4** (the left side of a ship or aircraft to someone who is aboard and facing the bow or nose)

interface, **port.n.5** ((computer science) computer circuit consisting of the hardware and associated circuitry that links one device with another (especially a computer and a hard disk drive or other peripherals))

Word Sense Disambiguation

A classification problem:

General Form:

$f(\text{sent_tokens}, (\text{target_index}, \text{lemma}, \text{POS})) \rightarrow \text{word_sense}$

port.n.1
port.n.2
port.n.3,
port.n.4
port.n.5

He walked along the **port** next to the steamer.

Word Sense Disambiguation

A classification problem:

General Form:

$$f(\text{sent_tokens}, (\text{target_index}, \text{lemma}, \text{POS})) \rightarrow \text{word_sense}$$

Logistic Regression (or any discriminative classifier):

$$P_{\text{lemma}, \text{POS}}(\text{sense} = s \mid \text{features})$$

He walked along the **port** next to the steamer.

Word Sense Disambiguation

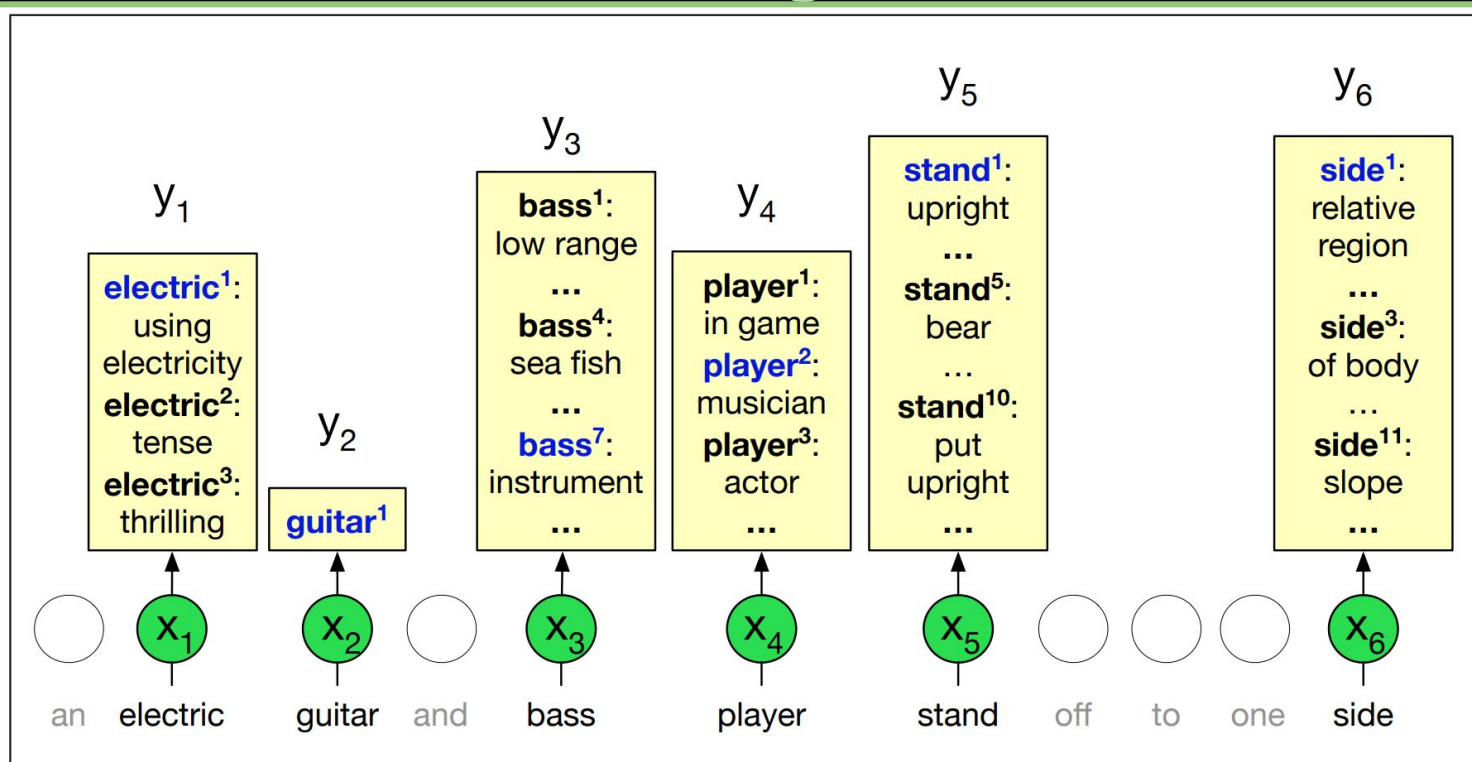


Figure 19.8 The all-words WSD task, mapping from input words (x) to WordNet senses (y). Only nouns, verbs, adjectives, and adverbs are mapped, and note that some words (like *guitar* in the example) only have one sense in WordNet. Figure inspired by [Chaplot and Salakhutdinov \(2018\)](#).

Distributional Hypothesis:

Wittgenstein, 1945: “*The meaning of a word is its use in the language*”

Distributional Hypothesis:

Wittgenstein, 1945: “*The meaning of a word is its use in the language*”

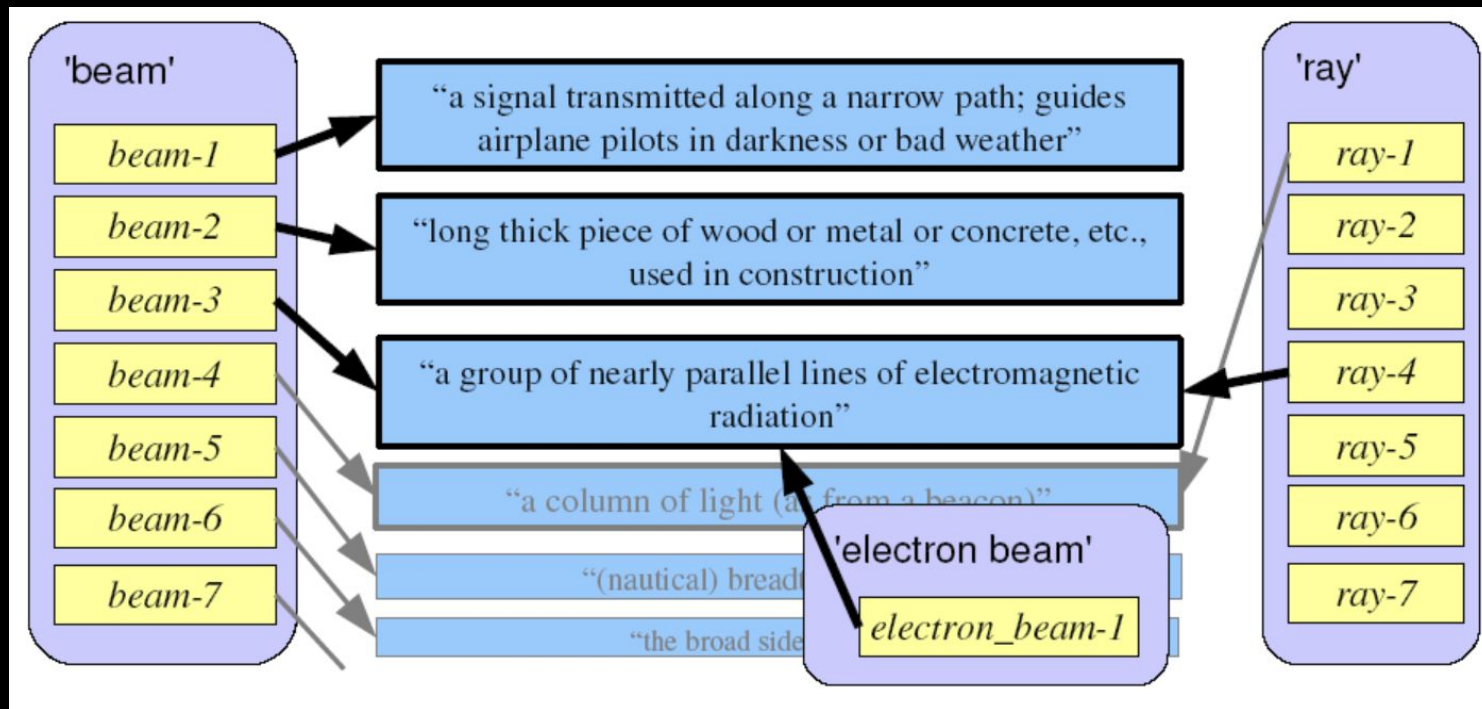
Distributional hypothesis -- A word's meaning is defined by all the different contexts it appears in (i.e. how it is “distributed” in natural language).

Firth, 1957: “*You shall know a word by the company it keeps*”

The nail hit the beam behind the wall.



Distributional Hypothesis



The nail hit the beam behind the wall.



Approaches to WSD

I.e. how to operationalize the distributional hypothesis.

1. Bag of words for context
E.g. multi-hot for any word in a defined “context”.
2. Surrounding window with positions
E.g. one-hot per position relative to word).
3. Lesk algorithm
E.g. compare context to sense definitions.
4. Selectors -- other *target* words that appear with same context
E.g. counts for any selector.
5. Contextual Embeddings
E.g. real valued vectors that “encode” the context (TBD).

Approaches to WSD

I.e. how to operationalize the distributional hypothesis.

1. Bag of words for context
E.g. multi-hot for any word in a defined “context”.
2. Surrounding window with positions
E.g. one-hot per position relative to word).
3. **Lesk algorithm**
E.g. compare context to sense definitions.
4. Selectors -- other *target* words that appear with same context
E.g. counts for any selector.
5. Contextual Embeddings
E.g. real valued vectors that “encode” the context (TBD).

Lesk Algorithm for WSD

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

best-sense \leftarrow most frequent sense for *word*

max-overlap $\leftarrow 0$

context \leftarrow set of words in *sentence*

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

signature \leftarrow set of words in the gloss and examples of *sense*

overlap \leftarrow COMPUTEOVERLAP(*signature*, *context*)

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

max-overlap \leftarrow *overlap*

best-sense \leftarrow *sense*

end

return(*best-sense*)

Figure 19.10 The 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.

Lesk Algorithm for WSD

- bank.n.1 (sloping land (especially the slope beside a body of water)) "they pulled the canoe up on the bank"; "he sat on the bank of the river and watched the currents"
- bank.n.2 (a financial institution that accepts deposits and channels the money into lending activities) "he cashed a check at the bank"; "that bank holds the mortgage on my home"

```
overlap  $\leftarrow$  COMPUTEOVERLAP(signature, context)  
if overlap > max-overlap then  
    max-overlap  $\leftarrow$  overlap  
    best-sense  $\leftarrow$  sense  
end  
return(best-sense)
```

The bank can guarantee deposits will cover future tuition costs, ...

- bank.n.1 (sloping land (especially the slope beside a body of water)) "they pulled the canoe up on the bank"; "he sat on the bank of the river and watched the currents"
- bank.n.2 (a financial institution that accepts deposits and channels the money into lending activities) "he cashed a check at the bank"; "that bank holds the mortgage on my home"
- ...
- bank.n.4 (an arrangement of similar objects in a row or in tiers) "he operated a bank of switches"
- ...
- bank.n.8 (a building in which the business of banking transacted) "the bank is on the corner of Nassau and Witherspoon"
- bank.n.9 (a flight maneuver; aircraft tips laterally about its longitudinal axis (especially in turning)) "the plane went into a steep bank"

end

return(*best-sense*)

The bank can guarantee deposits will cover future tuition costs, ...

Look-Algorithm for WSD

- **striker.n.1** (a forward on a soccer team)
- **striker.n.2** (someone receiving intensive training for a naval technical rating)
- **striker.n.3** (an employee on strike against an employer)
- **striker.n.4** (someone who hits) *"a hard hitter"; "a fine striker of the ball"; "blacksmiths are good hitters"*
- **striker.n.5** (the part of a mechanical device that strikes something)

overlap \leftarrow COMPUTEOVERLAP(*signature*, *context*)

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

max-overlap \leftarrow *overlap*

best-sense \leftarrow *sense*

end

return(*best-sense*)

He addressed the strikers at the rally.

Approaches to WSD

I.e. how to operationalize the distributional hypothesis.

1. Bag of words for context
E.g. multi-hot for any word in a defined “context”.
2. Surrounding window with positions
E.g. one-hot per position relative to word).
3. **Lesk algorithm**
E.g. compare context to sense definitions.
4. Selectors -- other *target* words that appear with same context
E.g. counts for any selector.
5. Contextual Embeddings
E.g. real valued vectors that “encode” the context (TBD).

Approaches to WSD

I.e. how to operationalize the distributional hypothesis.

1. Bag of words for context
E.g. multi-hot for any word in a defined “context”.
2. Surrounding window with positions
E.g. one-hot per position relative to word).
3. Lesk algorithm
E.g. compare context to sense definitions.
4. **Selectors -- other target words that appear with same context** *E.g. counts for any selector.*
5. Contextual Embeddings
E.g. real valued vectors that “encode” the context (TBD).

Selectors


... a word which can take the place of another given word within the same local context (Lin, 1997)

Original version: Local context defined by dependency parse

Selectors

... a word which can take the place of another given word within the same local context (Lin, 1997)

Original version: Local context defined by dependency parse



He addressed the strikers at the rally.

Selectors

... a word which can take the place of another given word within the same local context (Lin, 1997)

Original version: Local context defined by dependency parse (Lin, 1997)

Web version: Local context defined by lexical patterns matched on the Web (Schwartz, 2008).

*“He addressed the * at the rally.”*

Selectors

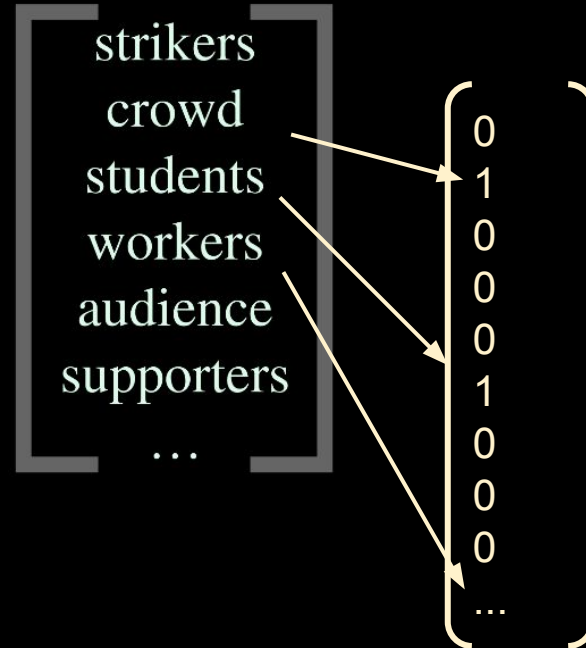
"He addressed the strikers at the rally."

strikers
crowd
students
workers
audience
supporters

...

Selectors

"He addressed the strikers at the rally."



Selectors

"He addressed the strikers at the rally."

he
man
owners
Mary
...

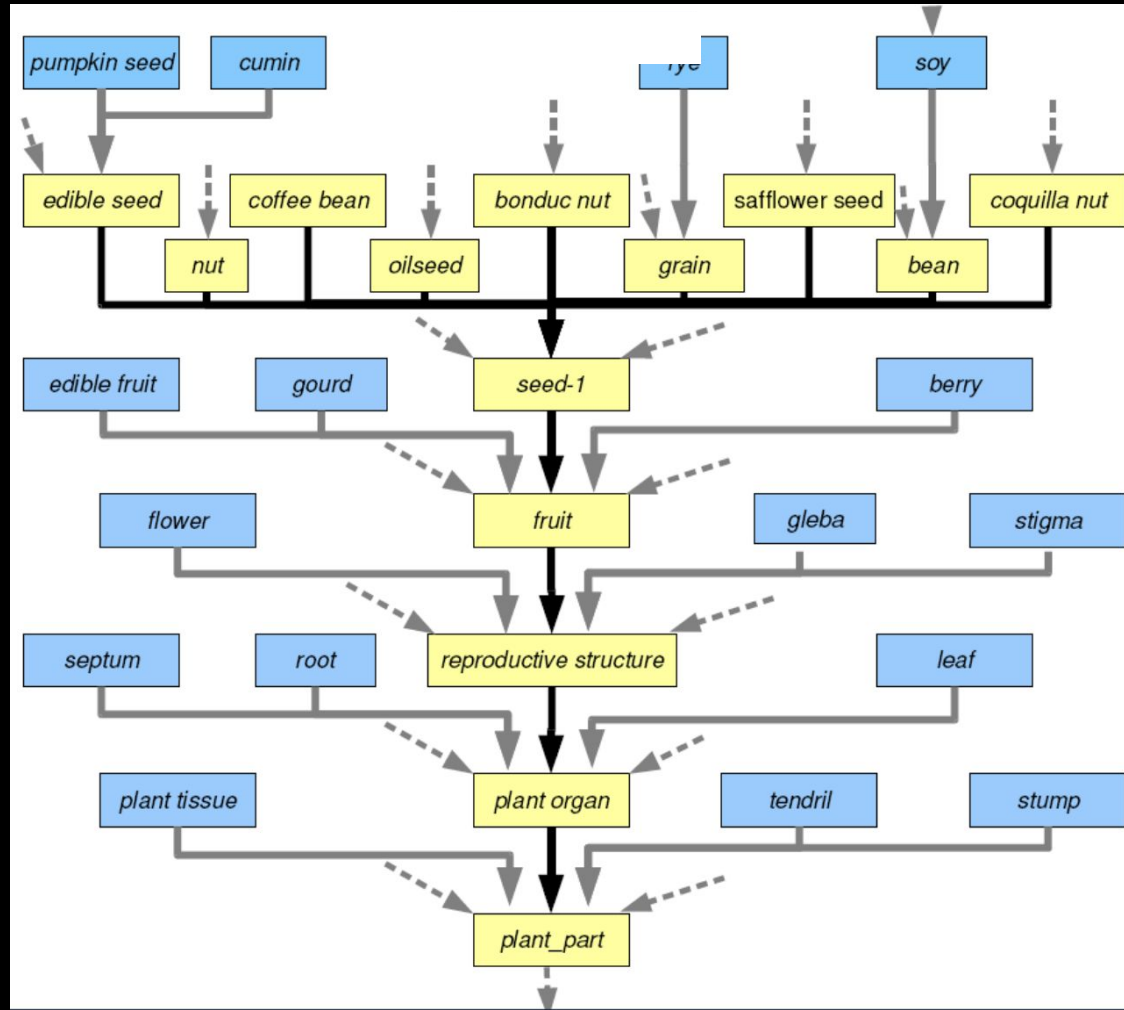
addressed
scolded
rallyied
kept
...

strikers
crowd
students
workers
audience
supporters
...

rally
protest
demonstration
work
stadium
...

Selectors

Leverages *hyponymy*:
concept1 <is-a> concept2



Why Are Selectors Effective?

Sets of selectors tend to vary extensively by word sense:

<i>bill-n.1</i>	<i>bill-n.2</i>	<i>bill-n.3</i>
bill	bill	market
it	staff	system
legislation	system	paper
system	money	note
program	time	bill
law	it	bond
plan	tax	stock
you	work	debt
measure	rent	rate
project	tuition	report

<i>occur-v.1</i>	<i>occur-v.2</i>	<i>occur-v.3</i>
be	go	go
happen	get	look
occur	Come	break
go	have	remove
take	try	find
work	lead	get
come	listen	place
see	work	keep
have	be	stick
change	belong	stop

- Polls show wide, generalized support for some vague concept of service, but the **bill** now under discussion lacks any passionate public backing.
training set never contained: “but the _ now under”
- ... in his lecture, refers to the “startling experience which almost every person confesses, that particular passages of conversation and action have **occurred** to him in the same order before, whether dreaming or waking ...
small context is contradictory:
“action have occurred” => occur-v.1 (“to happen or take place”)
“occurred to him” => occur-v.2 (“to come to mind”)

<i>bill-n.1</i>	<i>bill-n.2</i>	<i>bill-n.3</i>
bill	bill	market
it	staff	system
legislation	system	paper
system	money	note
program	time	bill
law	it	bond
plan	tax	stock
you	work	debt
measure	rent	rate
project	tuition	report

<i>occur-v.1</i>	<i>occur-v.2</i>	<i>occur-v.3</i>
be	go	go
happen	get	look
occur	Come	break
go	have	remove
take	try	find
work	lead	get
come	listen	place
see	work	keep
have	be	stick
change	belong	stop

Supervised Selectors

	base	w/ sels	<i>mfs</i>	<i>tests</i>
noun	87.9	91.7	80.9	2559
verb	83.3	83.7	76.5	2292
both	85.7	87.9	78.8	4851

Accuracy over SemEval-2007: Task 17.

Vector Semantics

1. Word2vec
2. Topic Modeling - Latent Dirichlet Allocation (LDA)

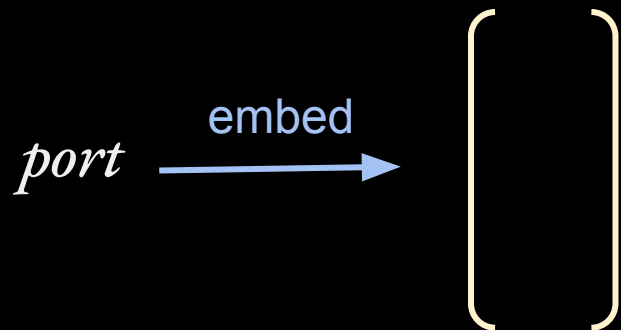
Objective

To embed: convert a token (or sequence) to a vector that **represents meaning**.

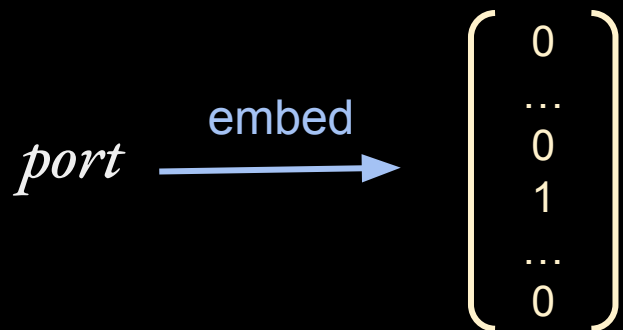
Objective

To embed: convert a token (or sequence) to a vector that represents meaning, or is useful to perform downstream NLP application.

Objective

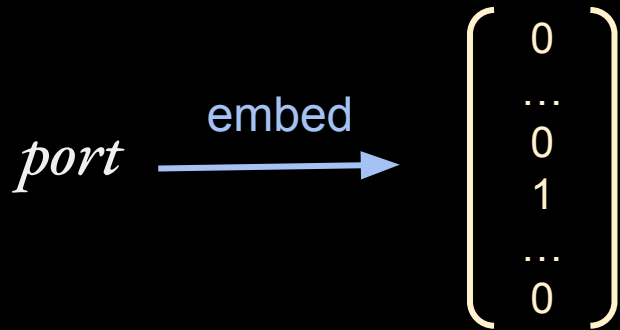


Objective



Objective

one-hot is sparse vector



Prefer dense vectors

- Less parameters (weights) for machine learning model.
- May generalize better implicitly.
- May capture synonyms

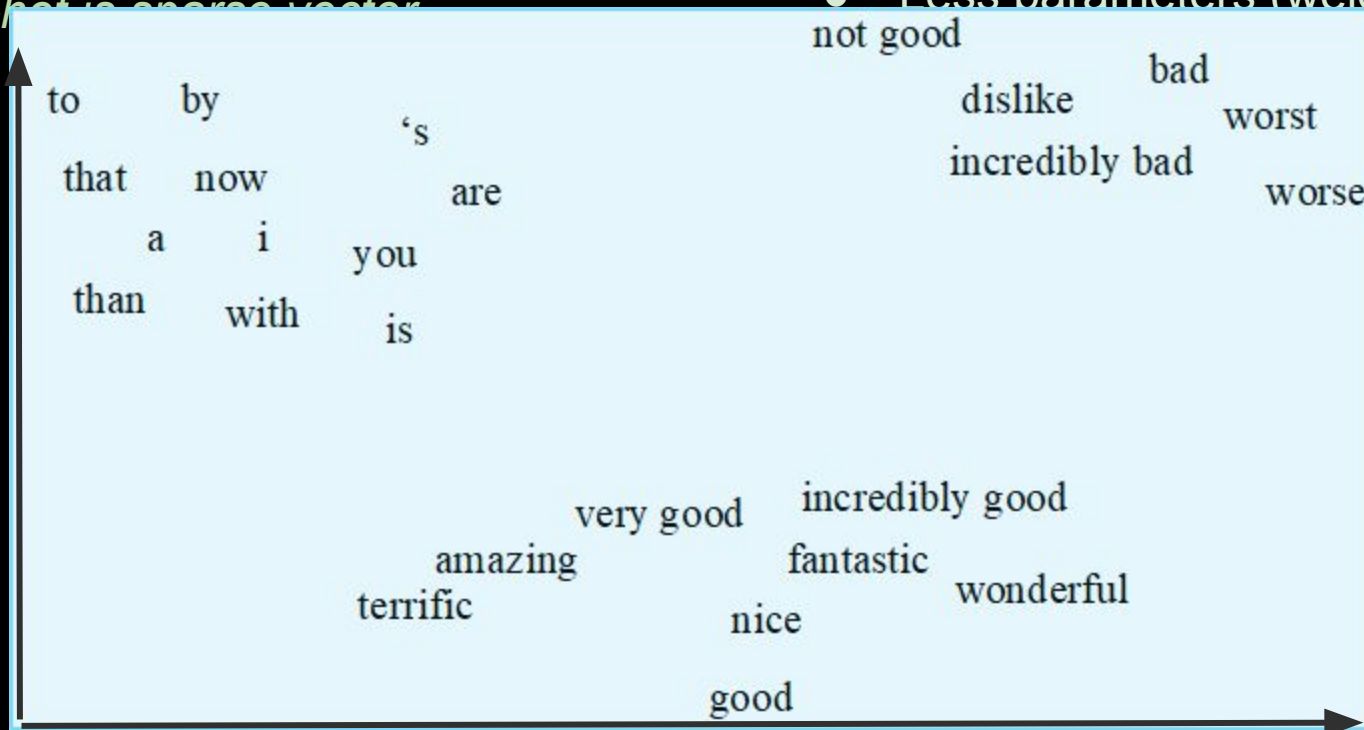
For deep learning, in practice, they work better. Why? Roughly, less parameters becomes increasingly important when you are learning multiple layers of weights rather than just a single layer.

Objective

Prefer dense vectors

- Less parameters (weights) for

one-hot is sparse vector

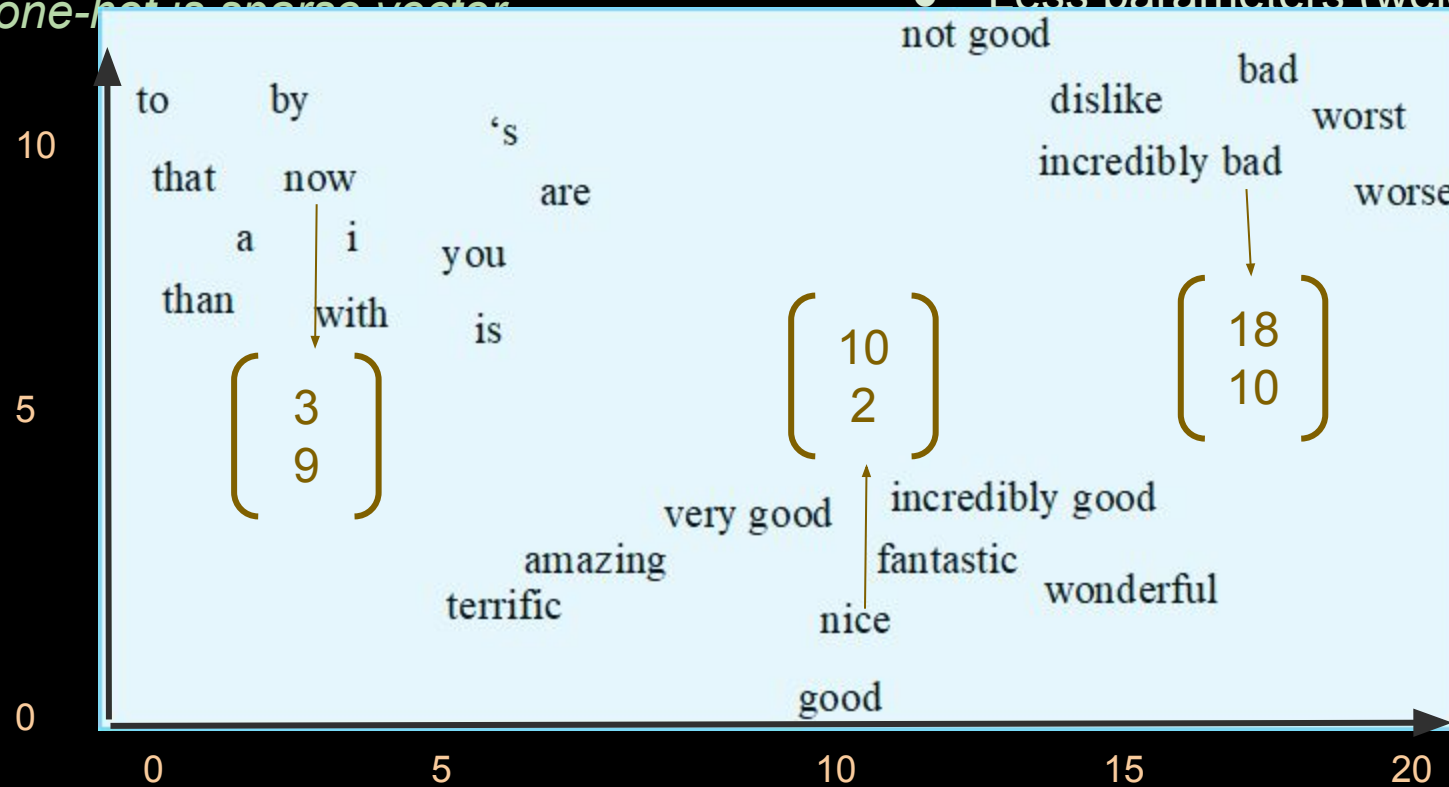


el.
implicitly.

ce, they work
rameters
nt when you are
ghts rather than

Objective

one-hot is sparse vector



Prefer dense vectors

- Less parameters (weights) for el.

implicitly.
S
ce, they work
rameters
nt when you are
ghts rather than

(Jurafsky, 2012)

Objective

To embed: convert a token (or sequence) to a vector that represents **meaning**.

Objective

To embed: convert a token (or sequence) to a vector that represents **meaning**.

Wittgenstein, 1945: “*The meaning of a word is its use in the language*”

Distributional hypothesis -- A word's meaning is defined by all the different contexts it appears in (i.e. how it is “distributed” in natural language).

Firth, 1957: “*You shall know a word by the company it keeps*”

Objective

To embed: convert a token (or sequence) to a vector that represents **meaning**.

Wittgenstein, 1945: “*The meaning of a word is its use in the language*”

Distributional hypothesis -- A word's meaning is defined by all the different contexts it appears in (i.e. how it is “distributed” in natural language).

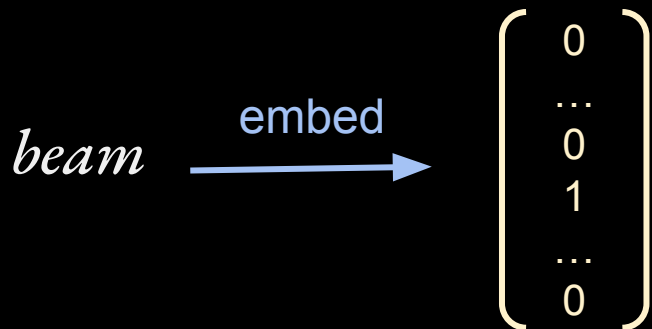
Firth, 1957: “*You shall know a word by the company it keeps*”

The nail hit the beam behind the wall.



Word Vectors

"one-hot encoding"

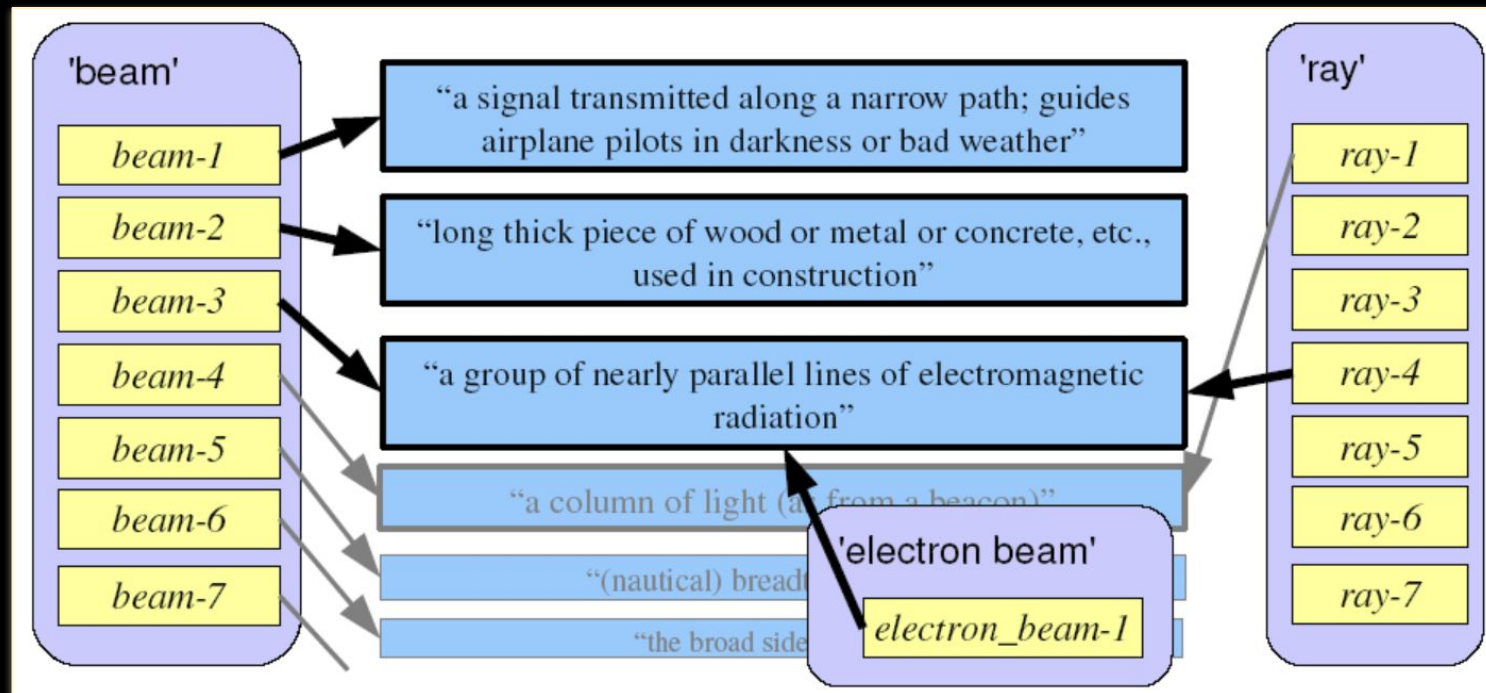


Prefer dense vectors

- Less parameters (weights) for machine learning model.
- May generalize better implicitly.
- May capture synonyms

For deep learning, in practice, they work better. Why? Roughly, less parameters becomes increasingly important when you are learning multiple layers of weights rather than just a single layer.

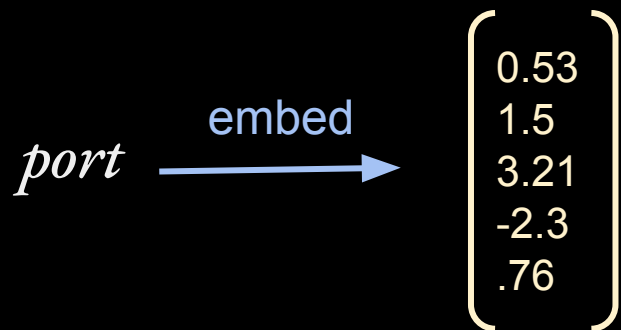
Distributional Hypothesis



The nail hit the beam behind the wall.

⏟ ⏟

Objective



Objective

