Lexical Semantics and Word Sense Disambiguation

CS224N // 18 November 2013

Spence Green

Stanford University

(with material from Bill MacCartney and Anoop Sarkar)

What is “computational semantics”?

Building and reasoning with **meaning representations**

Logic parsing	Natural language→logical form
Sentiment analysis	Classify (coarse) opinions
Semantic role labeling	Classify verb arguments
Word sense disambiguation	*bank* vs. *bank*

Three levels of meaning

1. Lexical Semantics
   - The meanings of individual words
2. Sentential / Compositional / Formal Semantics
   - How those meanings combine to make meanings for individual sentences or utterances
3. Discourse or Pragmatics
   - How those meanings combine with each other and with other facts about various kinds of context to make meanings for a text or discourse

(+ Dialog or Conversational Semantics)

Levels of meaning

Example: **short** poetic and religious text

Levels of meaning

Example: **short** poetic and religious text

Levels of meaning

Example: **short** poetic and religious text

```
It goes on.
```

Robert Frost on life

Levels of meaning

Example: **short** poetic and religious text

`It goes on.`	Robert Frost on life
إِنَّآ أَعْطَيْنٰكَ ٱلْكَوْثَرَ	108:1

*Indeed **we** gave **you** 'the abundance'*

Levels of meaning

Example: **short** poetic and religious text

`It goes on.`	Robert Frost on life
إِنَّآ أَعْطَيْنٰكَ ٱلْكَوْثَرَ	108:1
*Indeed **we** gave **you** 'the abundance'*	
`Jesus wept.`	John 11:35

Levels of meaning

Example: **short** poetic and religious text

`It goes on.`	Robert Frost on life
إِنَّآ أَعْطَيْنٰكَ ٱلْكَوْثَرَ	108:1
*Indeed **we** gave **you** 'the abundance'*	
`Jesus wept.`	John 11:35
My friend Freddy wept.	

Levels of meaning

Example: **short** poetic and religious text

`It goes on.`	Robert Frost on life
إِنَّآ أَعْطَيْنٰكَ ٱلْكَوْثَرَ	108:1
*Indeed **we** gave **you** 'the abundance'*	
`Jesus wept.`	John 11:35
My friend Freddy wept.	

Less than 140 characters!

The unit of meaning is a *sense*

- One word can have multiple meanings:
  - *Instead, a **bank** can hold the investments in a custodial account in the client's name.*
  - *But as agriculture burgeons on the east **bank**, the river will shrink even more.*
- We say that a **sense** is a representation of one aspect of the meaning of a word.
- Thus **bank** here has two senses
  - Bank^1:
  - Bank^2:

What is a sense?

(financial) **bank** vs. (river) **bank** is one extreme

What is a sense?

(financial) **bank** vs. (river) **bank** is one extreme

Other extreme:

What is a sense?

(financial) **bank** vs. (river) **bank** is one extreme

Other extreme:

1. **Plato** is on the shelf next to **Russell**.

What is a sense?

(financial) **bank** vs. (river) **bank** is one extreme

Other extreme:

1. **Plato** is on the shelf next to **Russell**.

2. **John** got a dent in his fender.

[Jackendoff, 2002]

Some more terminology

- Lemmas and wordforms
  - A **lexeme** is an abstract pairing of meaning and form
  - A **lemma** or **citation form** is the grammatical form that is used to represent a **lexeme**.
    - ***Carpet*** is the lemma for ***carpets***
    - ***Dormir*** is the lemma for ***duermes***
  - Specific surface forms *carpets, sung, duermes* are called **wordforms**
- The lemma *bank* has two **senses:**
  - *Instead, a **bank** can hold the investments in a custodial account in the client's name.*
  - *But as agriculture burgeons on the east **bank**, the river will shrink even more.*
- A **sense** is a discrete representation of one aspect of the meaning of a word

Tools of lexical semantics

Lexical relationships:

polysemy

homonymy

synonymy, antonymy, etc.

After Thanksgiving: argument structure, thematic roles, etc.

strange (adj.)

Define in terms of relationships:

strange (adj.)

Define in terms of relationships:

1. *weird, irregular, unusual* (synonyms)

strange (adj.)

Define in terms of relationships:

1. *weird, irregular, unusual* (synonyms)
2. *normal, regular, common* (antonyms)

strange (adj.)

Define in terms of relationships:

1. *weird, irregular, unusual* (synonyms)
2. *normal, regular, common* (antonyms)

Note: *Strangely*, *strangeness*, etc. all derived with morphological rules

Synonyms and antonyms

synonym: different orthography, same word sense

- *truck/lorry*
- *vomit/puke*
- *hack/program*

antonym: different orthography, opposite sense

- *tall/short*
- *rich/poor*
- *clean/dirty*

bank (n.)

bank (n.)

Polysemes and **Homonyms**:

1. a depository financial institution

bank (n.)

Polysemes and **Homonyms**:

1. a depository financial institution
2. heap of earth, e.g., beside a river (Homonym)

bank (n.)

Polysemes and **Homonyms**:

1. a depository financial institution
2. heap of earth, e.g., beside a river (Homonym)
3. objects arranged in tiers (polysemous w/ (2))

Polysemy vs. homonymy

Polyseme: same orthography, same pronunciation, different but related meanings (central origin)

Polysemy vs. homonymy

Polyseme: same orthography, same pronunciation, different but related meanings (central origin)

mouth of a river

mouth of an animal

Polysemy vs. homonymy

Polyseme: same orthography, same pronunciation, different but related meanings (central origin)

mouth of a river

mouth of an animal

Homonym: same orthography, same pronunciation, different, unrelated meanings

(financial) bank vs. (river) bank

More exotic: Troponyms

troponym: verb that indicates more precisely the manner of a more general verb

eat/nibble

run/sprint

defeat/vanquish

“One sort of structure universally acknowledged as part of lexical semantics is **taxonomic structure**....”

– Jackendoff (2002)

WordNet

Hierarchical lexical database organized by lexical semantic relations (Miller et al., 1990)

...an attempt to organize lexical information in terms of word meanings, rather than word forms. In that respect, WordNet resembles a thesaurus more than a dictionary.

sense: a discrete representation of one aspect of a word's meaning

WordNet

Current corpus size (periodic revisions):

POS	#
Noun	117,097
Adjective	22,141
Verb	11,488
Adverb	4,601

Has spawned VerbNet, *Net for other languages

WordNet

synset (synonym set): Set of near-synonyms

Ex. *bloke* → `{ chap, fellow, feller, fella, lad, gent, blighter, cuss}`

- What is the root?
- *bloke* → *entity* (hypernym)

WordNet

synset (synonym set): Set of near-synonyms

Ex. *bloke* → `{ chap, fellow, feller, fella, lad, gent, blighter, cuss}`

- What is the root?
- *bloke* → *entity* (hypernym)

Synset is the WordNet version of a “sense”

WordNet

`http://wordnetweb.princeton.edu/perl/webwn`

(Use this for the in-class exercise)

Supervised learning with WordNet

Supersense tagging: like NER for other POS categories

Taxonomy induction: automatic expansion of WordNet

Information retrieval, question answering, etc.

As features for many other tasks

In class today:

Word sense disambiguation with less supervision?

Semi-supervised learning: Bootstrapping

Bootstrapping: training on examples predicted in previous rounds

1. Label a few “seed” examples
2. Train a model
3. Classify some unlabeled examples
4. Select some unlabeled examples and extract features
5. Goto (2) until convergence

Semi-supervised learning: Bootstrapping

Bootstrapping pops up in many famous papers:

- Word sense disambiguation (Yarowsky, 1995)
- Named entity classification (Collins and Singer, 1999)
- Self training for parsing (McClosky et al., 2006)
- Pronominal coref (Bergsma and Lin, 2006)
- Many others!

Yarowsky Algorithm for WSD

Key insights:

1. **One sense per collocation**: Nearby words (*collocations*) provide strong and consistent clues to the sense of a target word.
2. **One sense per discourse**: The sense of a target word is highly consistent within any given *document*.

Yarowsky: One sense per collocation

Collocation	Frequency as **Aid**	Frequency as **Aide**
foreign	718	1
federal	297	0
western	146	0
provide	88	0
covert	26	0
oppose	13	0
future	9	0
similar	6	0
presidential	0	63
chief	0	40
longtime	0	26
aids-infected	0	2
sleepy	0	1
disaffected	0	1
indispensable	2	1
practical	2	0
squander	1	0

Yarowsky: One sense per discourse

The one-sense-per-discourse hypothesis:

Word	Senses	Accuracy	Applicblty
plant	living/factory	99.8 %	72.8 %
tank	vehicle/contnr	99.6 %	50.5 %
poach	steal/boil	100.0 %	44.4 %
palm	tree/hand	99.8 %	38.5 %
axes	grid/tools	100.0 %	35.5 %
sake	benefit/drink	100.0 %	33.7 %
bass	fish/music	100.0 %	58.8 %
space	volume/outer	99.2 %	67.7 %
motion	legal/physical	99.9 %	49.8 %
crane	bird/machine	100.0 %	49.1 %
Average		**99.8 %**	**50.1 %**

Example task: word sense disambiguation

303 unlabelled training examples:

- Full time should be served for each **sentence** .
- The Liberals inserted a **sentence** of 14 words which reads :
- They get a concurrent **sentence** with no additional time added to their sentence .
- The words tax relief appeared in every second **sentence** in the federal government's throne speech .

⋮

2 seed rules:

context: served sense 1

context: reads sense 2

Example task: word sense disambiguation

303 unlabelled training examples:

- Full time should be served for each **sentence** .
- The Liberals inserted a **sentence** of 14 words which reads :
- They get a concurrent **sentence** with no additional time added to their sentence .
- The words tax relief appeared in every second **sentence** in the federal government's throne speech .

⋮

2 seed rules:

context: served	sense 1
context: reads	sense 2

Example task: word sense disambiguation

303 unlabelled training examples:

- Full time should be served for each **sentence** .
- The Liberals inserted a **sentence** of 14 words which reads :
- They get a concurrent **sentence** with no additional time added to their sentence .
- The words tax relief appeared in every second **sentence** in the federal government's throne speech .

⋮

2 seed rules:

context: served	sense 1
context: reads	sense 2

Example task: word sense disambiguation

303 unlabelled training examples:

- Full time should be served for each **sentence** .
- The Liberals inserted a **sentence** of 14 words which reads :
- They get a concurrent **sentence** with no additional time added to their sentence .
- The words tax relief appeared in every second **sentence** in the federal government's throne speech .

⋮

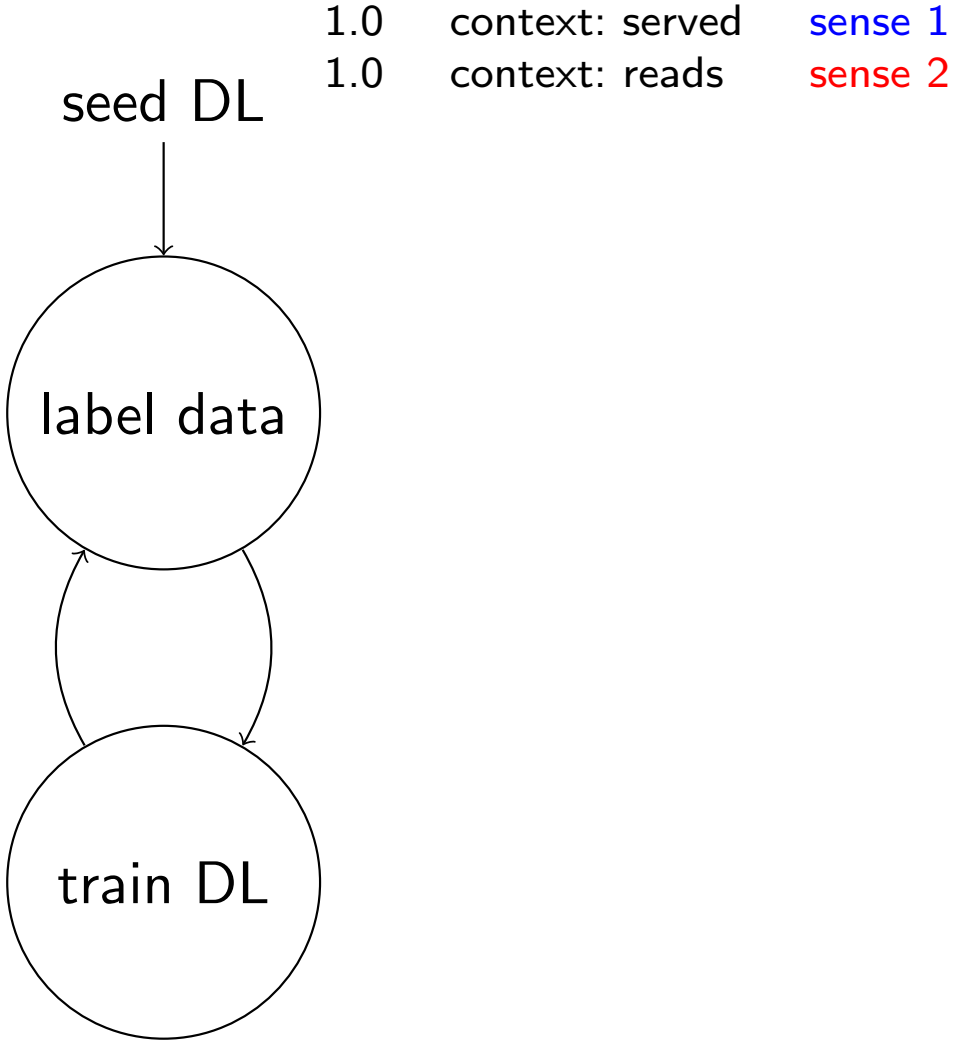
2 seed rules:

context: served	sense 1
context: reads	sense 2

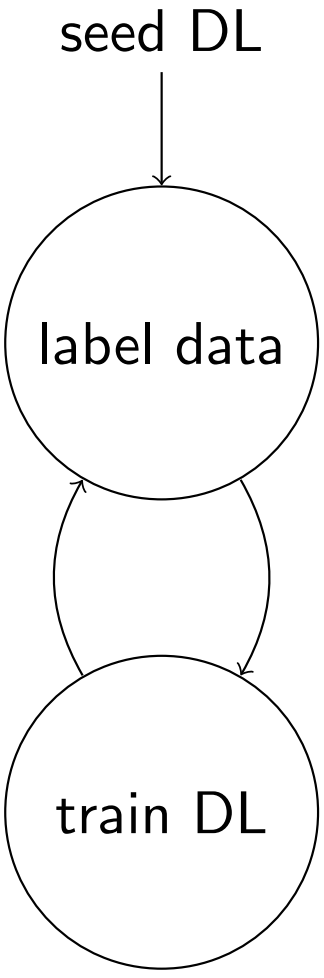
→ 76.99% accuracy on unseen test set

non-seeded accuracy (Daume, 2011)

Yarowsky algorithm (Yarowsky, 1995; Collins and Singer, 1999)



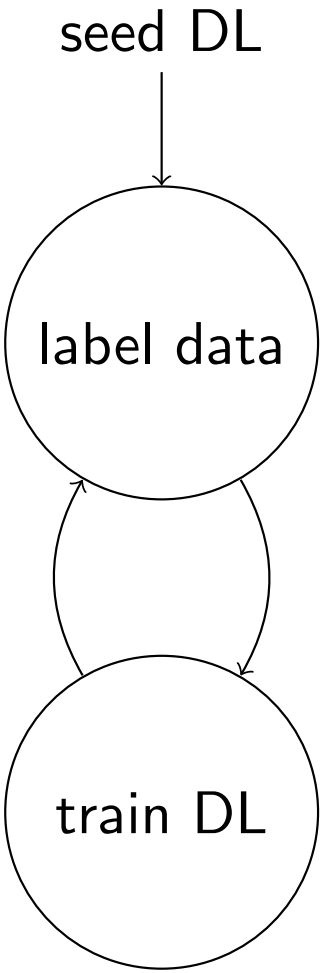
Yarowsky algorithm (Yarowsky, 1995; Collins and Singer, 1999)



1.0 context: served sense 1
1.0 context: reads sense 2

Full time should be [served] for each **sentence** .
The Liberals inserted a **sentence** of 14 words which [reads] :
The **sentence** for such an offence would be a term of imprisonment for one year .
Mr. Speaker , I have a question based on the very last **sentence** of the hon. member .
⋮

Yarowsky algorithm (Yarowsky, 1995; Collins and Singer, 1999)



1.0	context: served	sense 1
1.0	context: reads	sense 2

Full time should be [served] for each **sentence** .

The Liberals inserted a **sentence** of 14 words which [reads] :

The **sentence** for such an offence would be a term of imprisonment for one year .

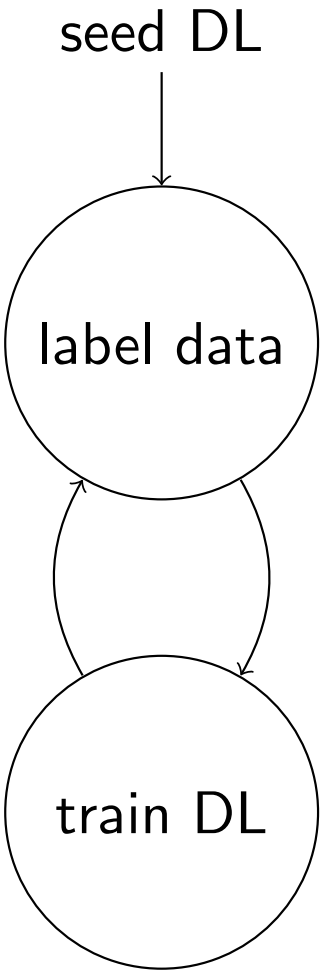
Mr. Speaker , I have a question based on the very last **sentence** of the hon. member .

⋮

1.0	context: served	sense 1
1.0	context: reads	sense 2
.976	context: serv*	sense 1
.976	context: read*	sense 2
.969	next word: reads	sense 2
.969	next word: read*	sense 2
.955	previous word: his	sense 1
.955	previous word: hi*	sense 1
.955	context: inmate	sense 1
.917	previous word: their	sense 1
.917	previous word: relevant	sense 2
.917	previous word: next	sense 2

⋮

Yarowsky algorithm (Yarowsky, 1995; Collins and Singer, 1999)



1.0	context: served	sense 1
1.0	context: reads	sense 2

Full time should be [served] for each **sentence** .

The Liberals inserted a **sentence** of 14 words which [reads] :

The **sentence** for such an offence would be a term of imprisonment for one year .

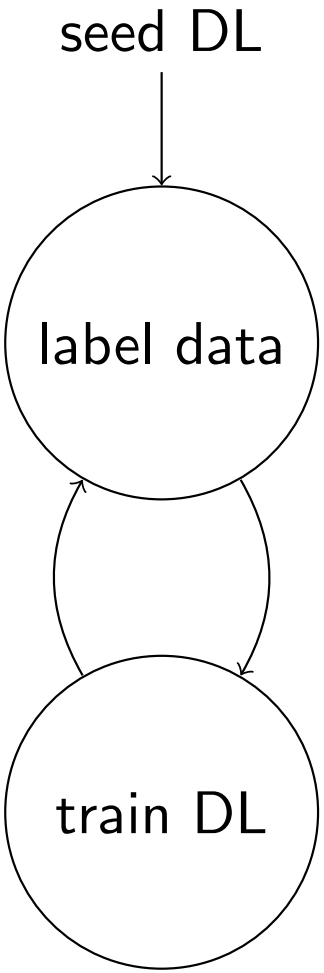
Mr. Speaker , I have a question based on the very last **sentence** of the hon. member .

⋮

1.0	context: served	sense 1
1.0	context: reads	sense 2
.976	context: serv*	sense 1
.976	context: read*	sense 2
.969	next word: reads	sense 2
.969	next word: read*	sense 2
.955	previous word: his	sense 1
.955	previous word: hi*	sense 1
.955	context: inmate	sense 1

threshold

Yarowsky algorithm (Yarowsky, 1995; Collins and Singer, 1999)



1.0 context: served sense 1
1.0 context: reads sense 2

Full time should be [served] for each **sentence** .
The Liberals inserted a **sentence** of 14 words which [reads] :
The **sentence** for such an offence would be a term of imprisonment for one year .
Mr. Speaker , I have a question based on the very last **sentence** of the hon. member .
⋮

1.0	context: served	sense 1
1.0	context: reads	sense 2
.976	context: serv*	sense 1
.976	context: read*	sense 2
.969	next word: reads	sense 2
.969	next word: read*	sense 2
.955	previous word: his	sense 1
.955	previous word: hi*	sense 1
.955	context: inmate	sense 1

threshold

Yarowsky Algorithm for WSD

Algorithm 1: The basic Yarowsky algorithm.

Require: training data X and a seed DL $\theta^{(0)}$

1: apply $\theta^{(0)}$ to X produce a labelling $Y^{(0)}$
2: **for** iteration t to maximum or convergence **do**
3: train a new DL θ on $Y^{(t)}$
4: apply θ to X, to produce $Y^{(t+1)}$
5: **end for**

Decision list classifier

Classify with the highest scoring rule that matches

Decision list classifier

Classify with the highest scoring rule that matches

Intuition: Use only the most reliable piece of evidence rather than the whole matching collocation set

Yarowsky (1995) features for "plant"

Final decision list for *plant* (abbreviated)		
LogL	Collocation	Sense
10.12	*plant* growth	⇒ A
9.68	car (within ±*k* words)	⇒ B
9.64	*plant* height	⇒ A
9.61	union (within ±*k* words)	⇒ B
9.54	equipment (within ±*k* words)	⇒ B
9.51	assembly *plant*	⇒ B
9.50	nuclear *plant*	⇒ B
9.31	flower (within ±*k* words)	⇒ A
9.24	job (within ±*k* words)	⇒ B
9.03	fruit (within ±*k* words)	⇒ A
9.02	*plant* species	⇒ A
...	...	

Interactive part

Your job:

1. Choose a word with multiple senses
2. `grep` it in a large corpus
3. Label a few examples of each sense
4. Write a few rules
5. Iterate with the Yarowsky algorithm

Code and data: `/afs/ir/class/cs224n/pa-lex/`

Report: `http://goo.gl/fUIvM2`