5525: Speech and Language Processing

Alan Ritter

(many slides from Greg Durrett)
Administrivia

- Piazza: link on the course website
- My office hours: Friday 4-5pm DL 595
- TA: Ashutosh Baheti; Office hours: Wednesday 1-2pm, DL 574
Course Requirements

- Probability
- Linear Algebra
- Calculus
- Programming / Python experience
- Prior exposure to machine learning very helpful but not required
Course Requirements

- Probability
- Linear Algebra
- Calculus
- Programming / Python experience
- Prior exposure to machine learning very helpful but not required

There will be a lot of math and programming!
Enrollment

- Homework 1 is out now (due August 30):
Homework 1 is out now (due August 30):

Please look at the assignment well before then
Homework 1 is out now (due August 30):

- Please look at the assignment well before then

- If this seems like it’ll be challenging for you, come and talk to me (this is smaller-scale than the later assignments, which are smaller-scale than the final project)
2 great textbooks for NLP

There will be assigned readings from both

Both freely available online

Speech and Language Processing (3rd ed. draft)
Dan Jurafsky and James H. Martin

Natural Language Processing
Jacob Eisenstein
What’s the goal of NLP?
What’s the goal of NLP?

- Be able to solve problems that require deep understanding of text
What’s the goal of NLP?

- Be able to solve problems that require deep understanding of text
- Example: dialogue systems
What’s the goal of NLP?

‣ Be able to solve problems that require deep understanding of text
‣ Example: dialogue systems
What’s the goal of NLP?

- Be able to solve problems that require deep understanding of text
- Example: dialogue systems

Siri, what’s the most valuable American company?
What’s the goal of NLP?

- Be able to solve problems that require deep understanding of text
- Example: dialogue systems

Siri, what’s the most valuable American company?

Apple
What’s the goal of NLP?

- Be able to solve problems that require deep understanding of text
- Example: dialogue systems

Siri, what’s the most valuable American company?

Apple

Who is its CEO?
What’s the goal of NLP?

- Be able to solve problems that require deep understanding of text
- Example: dialogue systems

Siri: what’s the most valuable American company?

Apple

Who is its CEO?

Tim Cook
What’s the goal of NLP?

- Be able to solve problems that require deep understanding of text
- Example: dialogue systems

recognize marketCap is the target value

Siri, what’s the most valuable American company?

Apple

Who is its CEO?

Tim Cook
What’s the goal of NLP?

- Be able to solve problems that require deep understanding of text
- Example: dialogue systems

Example:
- Dialogue systems

Siri, what’s the most valuable American company?
- Apple
- Who is its CEO?
- Tim Cook

recognize `marketCap` is the target value

recognize predicate
What’s the goal of NLP?

- Be able to solve problems that require deep understanding of text
- Example: dialogue systems

Siri, what’s the most valuable American company?

Apple

Who is its CEO?

Tim Cook
What’s the goal of NLP?

- Be able to solve problems that require deep understanding of text
- Example: dialogue systems

Siri, what’s the most valuable American company?

Apple

Who is its CEO?

Tim Cook

recognize marketCap is the target value

do computation

recognize predicate

resolve references
Automatic Summarization
Google Critic Ousted From Think Tank Funded by the Tech Giant

WASHINGTON — In the hours after European antitrust regulators levied a record $2.7 billion fine against Google in late June, an influential Washington think tank learned what can happen when a tech giant that shapes public policy debates with its enormous wealth is criticized.

But not long after one of New America’s scholars posted a statement on the think tank’s website praising the European Union’s penalty against Google, Mr. Schmidt, who had been chairman of New America until 2016, communicated his displeasure with the statement to the group’s president, Anne-Marie Slaughter, according to the scholar.

Ms. Slaughter told Mr. Lynn that “the time has come for Open Markets and New America to part ways,” according to an email from Ms. Slaughter to Mr. Lynn. The email suggested that the entire Open Markets team — nearly 10 full-time employees and unpaid fellows — would be exiled from New America.
One of New America’s writers posted a statement critical of Google. Eric Schmidt, Google’s CEO, was displeased.

The writer and his team were dismissed.
One of New America’s writers posted a statement critical of Google. Eric Schmidt, Google’s CEO, was displeased. The writer and his team were dismissed.
A critic of Google was recently dismissed from a think tank funded by the tech giant. The writer, one of New America’s scholars, posted a statement on the think tank’s website, criticizing Google’s recent $2.7 billion antitrust fine. The statement was posted on the think tank’s website and praised the European Union’s decision to penalize Google.

Eric Schmidt, Google’s CEO, was displeased with the writer’s statement. Schmidt had been a chairman of New America until 2016 and communicated his displeasure to the group’s president, Anne-Marie Slaughter. Slaughter told the writer, Mr. Lynn, that “the time has come for Open Markets and New America to part ways,” according to an email from Slaughter to Mr. Lynn. The email suggested that the entire Open Markets team — nearly 10 full-time employees and unpaid fellows — would be exiled from New America.
Google Critic Ousted From Think Tank Funded by the Tech Giant

WASHINGTON — In the hours after European antitrust regulators levied a record $2.7 billion fine against Google in late June, an influential Washington think tank learned what can happen when a tech giant that shapes public policy debates with its enormous wealth is criticized.

But not long after one of New America’s scholars posted a statement on the think tank’s website praising the European Union’s penalty against Google, Mr. Schmidt, who had been chairman of New America until 2016, communicated his displeasure with the statement to the group’s president, Anne-Marie Slaughter, according to the scholar.

Ms. Slaughter told Mr. Lynn that “the time has come for Open Markets and New America to part ways,” according to an email from Ms. Slaughter to Mr. Lynn. The email suggested that the entire Open Markets team — nearly 10 full-time employees and unpaid fellows — would be exiled from New America.

One of New America’s writers posted a statement critical of Google. Eric Schmidt, Google’s CEO, was displeased.

The writer and his team were dismissed.
Machine Translation

People's Daily, August 30, 2017
Machine Translation

People’s Daily, August 30, 2017

Translate

English  French  Spanish  Chinese - detected

特朗普偕家人在白宫阳台观看百年一遇日全食
Trump Pope family watch a hundred years a year in the White House balcony
Trump Pope family watch a hundred years a year in the White House balcony.
NLP Analysis Pipeline
NLP Analysis Pipeline

Text

_____________________________
_____________________________
_____________________________
_____________________________
_____________________________
NLP Analysis Pipeline

Text

Text Analysis
Syntactic parses
Coreference resolution
Entity disambiguation
Discourse analysis
NLP Analysis Pipeline

Text

Text Analysis
- Syntactic parses
- Coreference resolution
- Entity disambiguation
- Discourse analysis

Annotations
NLP Analysis Pipeline

Text

Text Analysis
Syntactic parses
Coreference resolution
Entity disambiguation
Discourse analysis

Annotations

Applications
NLP Analysis Pipeline

Text

Text Analysis
Syntactic parses
Coreference resolution
Entity disambiguation
Discourse analysis

Annotations

Applications
Summarize
NLP Analysis Pipeline

Text

Text Analysis
Syntactic parses
Coreference resolution
Entity disambiguation
Discourse analysis

Annotations

Applications
Summarize
Extract information
NLP Analysis Pipeline

Text

Text Analysis
- Syntactic parses
- Coreference resolution
- Entity disambiguation
- Discourse analysis

Annotations

Applications
- Summarize
- Extract information
- Answer questions
NLP Analysis Pipeline

Text

Text Analysis
Syntactic parses
Coreference resolution
Entity disambiguation
Discourse analysis

Annotations

Applications
Summarize
Extract information
Answer questions
Identify sentiment
NLP Analysis Pipeline

Text Analysis
- Syntactic parses
- Coreference resolution
- Entity disambiguation
- Discourse analysis

Annotations

Applications
- Summarize
- Extract information
- Answer questions
- Identify sentiment
- Translate
NLP Analysis Pipeline

Text

Text Analysis
Syntactic parses
Coreference resolution
Entity disambiguation
Discourse analysis

Annotations

Applications
Summarize
Extract information
Answer questions
Identify sentiment
Translate

- NLP is about building these pieces!
NLP is about building these pieces!

All of these components are modeled with statistical approaches trained with machine learning.
How do we represent language?
How do we represent language?
How do we represent language?

Text

Labels

the movie was good

+
How do we represent language?

**Text**

- the movie was good
- Beyoncé had one of the best videos of all time

**Labels**

- the movie was good
- Beyoncé had one of the best videos of all time
  - subjective
How do we represent language?

<table>
<thead>
<tr>
<th>Text</th>
<th>Labels</th>
<th>Sequences/tags</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>the movie was good</em> + subjective</td>
<td><em>PERSON</em> Tom Cruise stars in the new <em>WORK_OF_ART</em> Mission Impossible film</td>
</tr>
</tbody>
</table>
How do we represent language?

Text

Labels
the movie was good
Beyoncé had one of the best videos of all time subjective

Sequences/tags
PERSON
Tom Cruise stars in the new WORK_OF_ART Mission Impossible film

Trees
S
  VP
    NP
      VBZ
      NN
      PP
        NP
          PP
            I eat cake with icing
How do we represent language?

- **Labels**
  - "the movie was good"
  - "Beyoncé had one of the best videos of all time subjective"

- **Sequences/tags**
  - PERSON
    - Tom Cruise
  - WORK_OF_ART
    - Mission Impossible film

- **Trees**
  - "I eat cake with icing"
  - λx. flight(x) ∧ dest(x)=Miami
  - flights to Miami
How do we use these representations?

Text Analysis
Labels
Sequences
Trees
...
How do we use these representations?

Text

Text Analysis
Labels
Sequences
Trees
...

Applications
How do we use these representations?

Text

Text Analysis
- Labels
- Sequences
- Trees
...

Applications
- Extract syntactic features
How do we use these representations?

Text

Text Analysis
- Labels
- Sequences
- Trees
...

Applications
- Extract syntactic features
- Tree-structured neural networks
How do we use these representations?

Text

Text Analysis
- Labels
- Sequences
- Trees
- ...

Applications
- Extract syntactic features
- Tree-structured neural networks
- Tree transducers (for machine translation)
How do we use these representations?

Text

Text Analysis
Labels
Sequences
Trees
...

Applications
Extract syntactic features
Tree-structured neural networks
Tree transducers (for machine translation)
...

Applica ons
How do we use these representations?

Text Analysis
- Labels
- Sequences
- Trees
- ... end-to-end models

Applications
- Extract syntactic features
- Tree-structured neural networks
- Tree transducers (for machine translation)
- ...

Text
Main question: What representations do we need for language? What do we want to know about it?
Main question: What representations do we need for language? What do we want to know about it?

Boils down to: what ambiguities do we need to resolve?
Why is language hard?
(and how can we handle that?)
Language is Ambiguous!

- Hector Levesque (2011): “Winograd schema challenge” (named after Terry Winograd, the creator of SHRDLU)
Language is Ambiguous!

- Hector Levesque (2011): “Winograd schema challenge” (named after Terry Winograd, the creator of SHRDLU)

The city council refused the demonstrators a permit because they _____ violence
Language is Ambiguous!

- Hector Levesque (2011): “Winograd schema challenge” (named after Terry Winograd, the creator of SHRDLU)

The city council refused the demonstrators a permit because they ______ violence.
Language is Ambiguous!

- Hector Levesque (2011): “Winograd schema challenge” (named after Terry Winograd, the creator of SHRDLU)

The city council refused the demonstrators a permit because they ______ violence
Language is Ambiguous!

- Hector Levesque (2011): “Winograd schema challenge” (named after Terry Winograd, the creator of SHRDLU)

The city council refused the demonstrators a permit because they advocated violence.
Hector Levesque (2011): “Winograd schema challenge” (named after Terry Winograd, the creator of SHRDLU)

The city council refused the demonstrators a permit because:

- they advocated
- they feared
- they ______ violence
Hector Levesque (2011): “Winograd schema challenge” (named after Terry Winograd, the creator of SHRDLU)

The city council refused the demonstrators a permit because they ______ violence
they feared

they advocated
Language is Ambiguous!

- Hector Levesque (2011): “Winograd schema challenge” (named after Terry Winograd, the creator of SHRDLU)

  The city council refused the demonstrators a permit because they ______ violence

  - they advocated
  - they feared

- This is so complicated that it’s an AI challenge problem! (AI-complete)
Language is Ambiguous!

- Hector Levesque (2011): “Winograd schema challenge” (named after Terry Winograd, the creator of SHRDLU)

The city council refused the demonstrators a permit because they ______ violence.

- They advocated
- They feared

- This is so complicated that it’s an AI challenge problem! (AI-complete)
- Referential/semantic ambiguity
Language is Ambiguous!

slide credit: Dan Klein
Language is Ambiguous!

- Headlines
Language is Ambiguous!

- Headlines
  - Teacher Strikes Idle Kids
Language is Ambiguous!

- Headlines
  - Teacher Strikes Idle Kids
  - Hospitals Sued by 7 Foot Doctors
Language is Ambiguous!

- Headlines
  - Teacher Strikes Idle Kids
  - Hospitals Sued by 7 Foot Doctors
  - Ban on Nude Dancing on Governor’s Desk
Language is Ambiguous!

- Headlines
  - Teacher Strikes Idle Kids
  - Hospitals Sued by 7 Foot Doctors
  - Ban on Nude Dancing on Governor’s Desk
  - Iraqi Head Seeks Arms
Language is Ambiguous!

- Headlines
  - Teacher Strikes Idle Kids
  - Hospitals Sued by 7 Foot Doctors
  - Ban on Nude Dancing on Governor’s Desk
  - Iraqi Head Seeks Arms
  - Stolen Painting Found by Tree
Language is Ambiguous!

- Headlines
  - Teacher Strikes Idle Kids
  - Hospitals Sued by 7 Foot Doctors
  - Ban on Nude Dancing on Governor’s Desk
  - Iraqi Head Seeks Arms
  - Stolen Painting Found by Tree
  - Kids Make Nutritious Snacks

slide credit: Dan Klein
Language is Ambiguous!

- Headlines
  - Teacher Strikes Idle Kids
  - Hospitals Sued by 7 Foot Doctors
  - Ban on Nude Dancing on Governor’s Desk
  - Iraqi Head Seeks Arms
  - Stolen Painting Found by Tree
  - Kids Make Nutritious Snacks
  - Local HS Dropouts Cut in Half
Language is Ambiguous!

- Headlines
  - Teacher Strikes Idle Kids
  - Hospitals Sued by 7 Foot Doctors
  - Ban on Nude Dancing on Governor’s Desk
  - Iraqi Head Seeks Arms
  - Stolen Painting Found by Tree
  - Kids Make Nutritious Snacks
  - Local HS Dropouts Cut in Half

- Syntactic/semantic ambiguity: parsing needed to resolve these, but need context to figure out which parse is correct
Language is **Really** Ambiguous!

- There aren’t just one or two possibilities which are resolved pragmatically
Language is **Really** Ambiguous!

- There aren’t just one or two possibilities which are resolved pragmatically

*il fait vraiment beau*
Language is Really Ambiguous!

- There aren’t just one or two possibilities which are resolved pragmatically

  It is really nice out

  *il fait vraiment beau*
Language is **Really** Ambiguous!

- There aren’t just one or two possibilities which are resolved pragmatically

  It is really nice out

  *il fait vraiment beau* → It’s really nice
Language is **Really** Ambiguous!

- There aren’t just one or two possibilities which are resolved pragmatically

  *It is really nice out*

  *It’s really nice*

  *The weather is beautiful*

  *il fait vraiment beau*
Language is **Really** Ambiguous!

- There aren’t just one or two possibilities which are resolved pragmatically

  - *It is really nice out*
  - *It’s really nice*
  - *The weather is beautiful*
  - *It is really beautiful outside*

*il fait vraiment beau*
Language is **Really** Ambiguous!

- There aren’t just one or two possibilities which are resolved pragmatically

  *il fait vraiment beau*  
  
  - It is really nice out
  - It’s really nice
  - The weather is beautiful
  - It is really beautiful outside
  - He makes truly beautiful
Language is **Really** Ambiguous!

- There aren’t just one or two possibilities which are resolved pragmatically

- *il fait vraiment beau* →
  - It is really nice out
  - It’s really nice
  - The weather is beautiful
  - It is really beautiful outside
  - He makes truly beautiful
  - He makes truly boyfriend
Language is **Really** Ambiguous!

- There aren’t just one or two possibilities which are resolved pragmatically

  *il fait vraiment beau*  
  
  - It is really nice out
  - It’s really nice
  - The weather is beautiful
  - It is really beautiful outside
  - He makes truly beautiful
  - He makes truly boyfriend
  - It fact actually handsome
Language is **Really** Ambiguous!

- There aren’t just one or two possibilities which are resolved pragmatically

  *It is really nice out*
  *It’s really nice*
  *The weather is beautiful*
  *It is really beautiful outside*
  *He makes truly beautiful*
  *He makes truly boyfriend*
  *It fact actually handsome*

- Combinatorially many possibilities, many you won’t even register as ambiguities, but systems still have to resolve them
What do we need to understand language?

- Lots of data!

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>Cela constituerait une solution transitoire qui permettrait de conduire à terme à une charte à valeur contraignante.</th>
</tr>
</thead>
<tbody>
<tr>
<td>HUMAN</td>
<td>That would be an interim solution which would make it possible to work towards a binding charter in the long term.</td>
</tr>
<tr>
<td>1x DATA</td>
<td>[this] [constituerait] [assistance] [transitoire] [who] [permistrait] [licences] [to] [terme] [to] [a] [charter] [to] [value] [contraignante] [.]</td>
</tr>
<tr>
<td>10x DATA</td>
<td>[it] [would] [a solution] [transitional] [which] [would] [of] [lead] [to] [term] [to] [a] [charter] [to] [value] [binding] [.]</td>
</tr>
<tr>
<td>100x DATA</td>
<td>[this] [would be] [a transitional solution] [which would] [lead to] [a charter] [legally binding] [.]</td>
</tr>
<tr>
<td>1000x DATA</td>
<td>[that would be] [a transitional solution] [which would] [eventually lead to] [a binding charter] [.]</td>
</tr>
</tbody>
</table>
What do we need to understand language?

- World knowledge: have access to information beyond the training data
What do we need to understand language?

- World knowledge: have access to information beyond the training data

DOJ greenlights Disney - Fox merger
What do we need to understand language?

- World knowledge: have access to information beyond the training data

*DOJ greenlights Disney - Fox merger*
What do we need to understand language?

- World knowledge: have access to information beyond the training data

**DOJ greenlights Disney-Fox merger**

*Department of Justice*
What do we need to understand language?

- World knowledge: have access to information beyond the training data

**Diagram:**
- **DOJ** greenlights Disney-Fox merger
- Department of Justice
- metaphor; “approves”
What do we need to understand language?

- World knowledge: have access to information beyond the training data

Department of Justice

**DOJ** greenlights Disney - Fox merger

metaphor; “approves”

- What is a green light? How do we understand what “green lighting” does?
What do we need to understand language?

- Grounding: learn what fundamental concepts actually mean in a data-driven way
What do we need to understand language?

- Grounding: learn what fundamental concepts actually mean in a data-driven way.

**Question:** What object is right of [02]?

Golland et al. (2010)
What do we need to understand language?

- Grounding: learn what fundamental concepts actually mean in a data-driven way

**Question:** What object is right of 02?

![Diagram of objects](Golland-et-al-2010)

**Golland et al. (2010)**

![Hue distribution](McMahan-and-Stone-2015)

**McMahan and Stone (2015)**
What do we need to understand language?

- Linguistic structure
What do we need to understand language?

- Linguistic structure
- ...but computers probably won’t understand language the same way humans do

Centering Theory
Grosz et al. (1995)
What do we need to understand language?

- Linguistic structure
- ...but computers probably won’t understand language the same way humans do
- However, linguistics tells us what phenomena we need to be able to deal with and gives us hints about how language works

Centering Theory
Grosz et al. (1995)
What do we need to understand language?

- Linguistic structure
- ...but computers probably won’t understand language the same way humans do
- However, linguistics tells us what phenomena we need to be able to deal with and gives us hints about how language works

a. John has been having a lot of trouble arranging his vacation.

b. He cannot find anyone to take over his responsibilities. (he = John)

   $C_b = \text{John}; C_f = \{\text{John}\}$

c. He called up Mike yesterday to work out a plan. (he = John)

   $C_b = \text{John}; C_f = \{\text{John, Mike}\}$ (CONTINUE)

d. Mike has annoyed him a lot recently.

   $C_b = \text{John}; C_f = \{\text{Mike, John}\}$ (RETAIN)

e. He called John at 5 AM on Friday last week. (he = Mike)

   $C_b = \text{Mike}; C_f = \{\text{Mike, John}\}$ (SHIFT)
What techniques do we use?
(to combine data, knowledge, linguistics, etc.)
A brief history of (modern) NLP

1980  1990  2000  2010  2018
A brief history of (modern) NLP

“AI winter”, rule-based, expert systems
A brief history of (modern) NLP

- Early stat MT work at IBM
- "AI winter" rule-based, expert systems
A brief history of (modern) NLP

- "AI winter" rule-based, expert systems
- Earliest stat MT work at IBM
- Penn treebank
  
  \[
  S \\
  \downarrow \\
  NP \quad VP
  \]

- 1980
- 1990
- 2000
- 2010
- 2018
A brief history of (modern) NLP

- **1980**: "AI winter" rule-based, expert systems
- **1990**: Earliest stat MT work at IBM
- **2000**: Penn treebank (S \( \rightarrow \) NP, VP)
- **2000**: Ratnaparkhi tagger (NNP, VBZ)
A brief history of (modern) NLP

- 1980: "AI winter", rule-based, expert systems
- 1990: Early statistical MT work at IBM
- 1990: Pennsylvania treebank
- 2000: Ratnaparkhi tagger (NNP VBZ)
- 2000: Collins vs. Charniak parsers

1980 1990 2000 2010 2018
A brief history of (modern) NLP

- "AI winter" rule-based, expert systems
- Earliest stat MT work at IBM
- Penn treebank
- Ratnaparkhi tagger
- Collins vs. Charniak parsers

Sup: SVMs, CRFs, NER, Sentiment
A brief history of (modern) NLP

- **1980**: Earliest statistical MT work at IBM
- **1990**: “AI winter” rule-based, expert systems
- **1990**: Penn treebank
- **2000**: Ratnaparkhi tagger
- **2000**: Collins vs. Charniak parsers
- **2010**: Unsup: topic models, grammar induction
- **2018**: Sup: SVMs, CRFs, NER, Sentiment
A brief history of (modern) NLP

1980: “AI winter”
   - Rule-based, expert systems
   - Earliest statistical MT work at IBM

1990: Penn treebank
   - S
     - NP
     - VP

2000: Collins vs. Charniak parsers
   - Ratnaparkhi tagger
     - NNP
     - VBZ

2010: Unsup: topic models, grammar induction
   - Supervised: SVMs, CRFs, NER, Sentiment

2018: Semi-sup, structured prediction

Supervised Learning

Unsupervised Learning

Semi-supervised Learning
A brief history of (modern) NLP

- **1980**: Earliest statistical MT work at IBM
- **1990**: "AI winter" rule-based, expert systems
- **1990**: Ratnaparkhi tagger
- **2000**: Penn treebank
- **2000**: Collins vs. Charniak parsers
- **2000**: Unsup: topic models, grammar induction
- **2000**: Sup: SVMs, CRFs, NER, Sentiment
- **2000**: Semi-sup, structured prediction
- **2000**: Neural

**Timeline**
- 1980
- 1990
- 2000
- 2010
- 2018
A brief history of (modern) NLP

- **1980**: “AI winter”
  - rule-based, expert systems
  - earliest stat MT work at IBM

- **1990**: Penn treebank
  - S
  - NP
  - VP

- **2000**: Collins vs. Charniak parsers
  - Ratnaparkhi tagger
  - NNP
  - VBZ

- **2010**: Unsup: topic models, grammar induction
  - Sup: SVMs, CRFs, NER, Sentiment
  - Semi-sup, structured prediction

- **2018**: Neural
Structured Prediction

“Learning a Part-of-Speech Tagger from Two Hours of Annotation”
Garrette and Baldridge (2013)
All of these techniques are data-driven! Some data is naturally occurring, but may need to label.
Structured Prediction

- All of these techniques are data-driven! Some data is naturally occurring, but may need to label.
- Supervised techniques work well on very little data.

“Learning a Part-of-Speech Tagger from Two Hours of Annotation”
Garrette and Baldridge (2013)
Structured Prediction

- All of these techniques are data-driven! Some data is naturally occurring, but may need to label.
- Supervised techniques work well on very little data.

“Learning a Part-of-Speech Tagger from Two Hours of Annotation”
Garrette and Baldridge (2013)
Structured Prediction

- All of these techniques are data-driven! Some data is naturally occurring, but may need to label
- Supervised techniques work well on very little data

“Learning a Part-of-Speech Tagger from Two Hours of Annotation”
Garrette and Baldridge (2013)
Structured Prediction

- All of these techniques are data-driven! Some data is naturally occurring, but may need to label.
- Supervised techniques work well on very little data.

“Learning a Part-of-Speech Tagger from Two Hours of Annotation”
Garrette and Baldridge (2013)
Structured Prediction

- All of these techniques are data-driven! Some data is naturally occurring, but may need to be labeled.
- Supervised techniques work well on very little data.

“Learning a Part-of-Speech Tagger from Two Hours of Annotation”
Garrette and Baldridge (2013)
Structured Prediction

- All of these techniques are data-driven! Some data is naturally occurring, but may need to be labeled.
- Supervised techniques work well on very little data.
- Even neural nets can do pretty well!

“Learning a Part-of-Speech Tagger from Two Hours of Annotation”
Garrette and Baldridge (2013)
Less Manual Structure?

(a) example word alignment

(b) example phrase alignment

DeNero et al. (2008)

Bahdanau et al. (2014)
Does manual structure have a place?
Does manual structure have a place?

- Neural nets don’t always work out of domain!
Does manual structure have a place?

- Neural nets don’t always work out of domain!

- Coreference: rule-based systems are still about as good as deep learning out-of-domain
Does manual structure have a place?

- Neural nets don’t always work out of domain!
- Coreference: rule-based systems are still about as good as deep learning out-of-domain

<table>
<thead>
<tr>
<th></th>
<th>CoNLL</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Avg. F₁</td>
<td></td>
</tr>
<tr>
<td><strong>Newswire</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rule-based</td>
<td>55.60</td>
<td></td>
</tr>
<tr>
<td>berkeley</td>
<td>61.24</td>
<td></td>
</tr>
<tr>
<td>cort</td>
<td>63.37</td>
<td></td>
</tr>
<tr>
<td>deep-coref [conll]</td>
<td>65.39</td>
<td></td>
</tr>
<tr>
<td>deep-coref [lea]</td>
<td>65.60</td>
<td></td>
</tr>
<tr>
<td><strong>Wikipedia</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rule-based</td>
<td>51.77</td>
<td></td>
</tr>
<tr>
<td>berkeley</td>
<td>51.01</td>
<td></td>
</tr>
<tr>
<td>cort</td>
<td>49.94</td>
<td></td>
</tr>
<tr>
<td>deep-coref [conll]</td>
<td>52.65</td>
<td></td>
</tr>
<tr>
<td>deep-coref [lea]</td>
<td>53.14</td>
<td></td>
</tr>
<tr>
<td>deep-coref⁻</td>
<td>51.01</td>
<td></td>
</tr>
</tbody>
</table>

Moosavi and Strube (2017)
Does manual structure have a place?

- Neural nets don’t always work out of domain!
- Coreference: rule-based systems are still about as good as deep learning out-of-domain
- LORELEI: transition point below which phrase-based systems are better

<table>
<thead>
<tr>
<th></th>
<th>CoNLL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Avg. F1</td>
</tr>
<tr>
<td>Newswire</td>
<td></td>
</tr>
<tr>
<td>rule-based</td>
<td>55.60</td>
</tr>
<tr>
<td>berkeley</td>
<td>61.24</td>
</tr>
<tr>
<td>cort</td>
<td>63.37</td>
</tr>
<tr>
<td>deep-coref [conll]</td>
<td>65.39</td>
</tr>
<tr>
<td>deep-coref [lea]</td>
<td>65.60</td>
</tr>
<tr>
<td>Wikipedia</td>
<td></td>
</tr>
<tr>
<td>rule-based</td>
<td>51.77</td>
</tr>
<tr>
<td>berkeley</td>
<td>51.01</td>
</tr>
<tr>
<td>cort</td>
<td>49.94</td>
</tr>
<tr>
<td>deep-coref [conll]</td>
<td>52.65</td>
</tr>
<tr>
<td>deep-coref [lea]</td>
<td>53.14</td>
</tr>
<tr>
<td>deep-coref−</td>
<td>51.01</td>
</tr>
</tbody>
</table>

Moosavi and Strube (2017)
Does manual structure have a place?

- Neural nets don’t always work out of domain!
- Coreference: rule-based systems are still about as good as deep learning out-of-domain
- LORELEI: transition point below which phrase-based systems are better
- Why is this? Inductive bias!

<table>
<thead>
<tr>
<th></th>
<th>CoNLL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Avg. F₁</strong></td>
</tr>
<tr>
<td>Newswire</td>
<td></td>
</tr>
<tr>
<td>rule-based</td>
<td>55.60</td>
</tr>
<tr>
<td>berkeley</td>
<td>61.24</td>
</tr>
<tr>
<td>cort</td>
<td>63.37</td>
</tr>
<tr>
<td>deep-coref [conll]</td>
<td>65.39</td>
</tr>
<tr>
<td>deep-coref [lea]</td>
<td>65.60</td>
</tr>
<tr>
<td>Wikipedia</td>
<td></td>
</tr>
<tr>
<td>rule-based</td>
<td>51.77</td>
</tr>
<tr>
<td>berkeley</td>
<td>51.01</td>
</tr>
<tr>
<td>cort</td>
<td>49.94</td>
</tr>
<tr>
<td>deep-coref [conll]</td>
<td>52.65</td>
</tr>
<tr>
<td>deep-coref [lea]</td>
<td>53.14</td>
</tr>
<tr>
<td>deep-coref^-</td>
<td>51.01</td>
</tr>
</tbody>
</table>

Moosavi and Strube (2017)
Does manual structure have a place?

- Neural nets don’t always work out of domain!
- Coreference: rule-based systems are still about as good as deep learning out-of-domain
- LORELEI: transition point below which phrase-based systems are better
- Why is this? Inductive bias!
- Can multi-task learning help?

<table>
<thead>
<tr>
<th></th>
<th>CoNLL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Avg. F₁</td>
</tr>
<tr>
<td><strong>Newswire</strong></td>
<td></td>
</tr>
<tr>
<td>rule-based</td>
<td>55.60</td>
</tr>
<tr>
<td>berkeley</td>
<td>61.24</td>
</tr>
<tr>
<td>cort</td>
<td>63.37</td>
</tr>
<tr>
<td>deep-coref [conll]</td>
<td>65.39</td>
</tr>
<tr>
<td>deep-coref [lea]</td>
<td>65.60</td>
</tr>
<tr>
<td><strong>Wikipedia</strong></td>
<td></td>
</tr>
<tr>
<td>rule-based</td>
<td>51.77</td>
</tr>
<tr>
<td>berkeley</td>
<td>51.01</td>
</tr>
<tr>
<td>cort</td>
<td>49.94</td>
</tr>
<tr>
<td>deep-coref [conll]</td>
<td>52.65</td>
</tr>
<tr>
<td>deep-coref [lea]</td>
<td>53.14</td>
</tr>
<tr>
<td>deep-coref</td>
<td>51.01</td>
</tr>
</tbody>
</table>

Moosavi and Strube (2017)
Does manual structure have a place?

Translate

Trump Pope family watch a hundred years a year in the White House balcony
Does manual structure have a place?

Trump, Pope, family watch a hundred years a year in the White House balcony.

- Maybe manual structure would help...
Where are we?
Where are we?

- NLP consists of: analyzing and building representations for text, solving problems involving text
Where are we?

- NLP consists of: analyzing and building representations for text, solving problems involving text

- These problems are hard because language is ambiguous, requires drawing on data, knowledge, and linguistics to solve
Where are we?

- NLP consists of: analyzing and building representations for text, solving problems involving text
- These problems are hard because language is ambiguous, requires drawing on data, knowledge, and linguistics to solve
- Knowing which techniques use requires understanding dataset size, problem complexity, and a lot of tricks!
NLP consists of: analyzing and building representations for text, solving problems involving text.

These problems are hard because language is ambiguous, requires drawing on data, knowledge, and linguistics to solve.

Knowing which techniques use requires understanding dataset size, problem complexity, and a lot of tricks!

NLP encompasses all of these things.
NLP vs. Computational Linguistics

- NLP: build systems that deal with language data
- CL: use computational tools to study language

Hamilton et al. (2016)
NLP vs. Computational Linguistics

- NLP: build systems that deal with language data
- CL: use computational tools to study language

Hamilton et al. (2016)
NLP vs. Computational Linguistics

- Computational tools for other purposes: literary theory, political science...
NLP vs. Computational Linguistics

- Computational tools for other purposes: literary theory, political science...

Bamman, O’Connor, Smith (2013)
### Outline of the Course

<table>
<thead>
<tr>
<th>Topics</th>
<th>Readings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction [4pp]</td>
<td></td>
</tr>
<tr>
<td>Binary classification</td>
<td>JM 6.1-6.3</td>
</tr>
<tr>
<td>Multiclass classification</td>
<td>JM 7, Structured SVM secs 1-2</td>
</tr>
<tr>
<td>Sequence models I: HMMs</td>
<td>JM 9, JM 10.4, Manning POS</td>
</tr>
<tr>
<td>Sequence models II: CRFs</td>
<td>Sutton CRFs 2.3, 2.6.1, Wallach CRFs tutorial, Illinois NER</td>
</tr>
<tr>
<td>Neural Nets II: FFNNs</td>
<td>Goldberg 1-4, 6, NLP with FFNNs, DANs</td>
</tr>
<tr>
<td>Neural Nets II: NN impl / word embeddings</td>
<td>Goldberg 5, word2vec, GloVe, Dropout</td>
</tr>
<tr>
<td>Neural Nets III: RNN and CNN encoders</td>
<td>Goldberg 9-11, Kim</td>
</tr>
<tr>
<td>Neural Nets IV: Neural CRFs</td>
<td>Collobert and Weston, Neural NER, Neural CRF parsing</td>
</tr>
<tr>
<td>Trees II: Dependency I</td>
<td>JM 14.1-14.4, Huang 1-2</td>
</tr>
<tr>
<td>Trees III: Dependency II</td>
<td>Parsley, Huang 2</td>
</tr>
<tr>
<td>Semantics I</td>
<td></td>
</tr>
<tr>
<td>Semantics II / Seq2seq I</td>
<td></td>
</tr>
<tr>
<td>Seq2seq II: Beam search, attention</td>
<td>Seq2seq, Attention, Luong Attention</td>
</tr>
<tr>
<td>Information Extraction / SRL</td>
<td>Distant supervision, RL for slot filling, TextRunner, ReVerb, NELL</td>
</tr>
<tr>
<td>Discourse and Coreference</td>
<td></td>
</tr>
<tr>
<td>Machine Translation I: Phrase-based</td>
<td>HMM alignment, Pharaoh</td>
</tr>
<tr>
<td>Machine Translation II: Neural</td>
<td></td>
</tr>
<tr>
<td>Applications I: Reading comprehension / MemNets</td>
<td>E2E Memory Networks, CBT, SQuAD, BiDAF</td>
</tr>
<tr>
<td>Applications II: Language grounding</td>
<td></td>
</tr>
<tr>
<td>Applications III: Summarization</td>
<td>MMR, Gillick, Sentence compression, SummaRUNNER, Pointer</td>
</tr>
<tr>
<td>Applications IV: Dialogue</td>
<td>RNN chatbots, Diversity, Goal-oriented, Latent Intention, QA-as-dialogue</td>
</tr>
<tr>
<td>Unsupervised Learning</td>
<td></td>
</tr>
<tr>
<td>NO CLASS (Thanksgiving)</td>
<td></td>
</tr>
<tr>
<td>Multilinguality and morphology</td>
<td></td>
</tr>
<tr>
<td>Wrapup</td>
<td></td>
</tr>
</tbody>
</table>
### Outline of the Course

<table>
<thead>
<tr>
<th>Topics</th>
<th>Readings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction [4pp]</td>
<td></td>
</tr>
<tr>
<td>Binary classification</td>
<td>JM 6.1-6.3</td>
</tr>
<tr>
<td>Multiclass classification</td>
<td>JM 7, Structured SVM secs 1-2</td>
</tr>
<tr>
<td>Sequence models I: HMMs</td>
<td>JM 9, JM 10.4, Manning POS</td>
</tr>
<tr>
<td>Sequence models II: CRFs</td>
<td>Sutton CRFs 2.3, 2.6.1, Wallach CRFs tutorial, Illinois NER</td>
</tr>
<tr>
<td>Neural Nets I: FFNNs</td>
<td>Goldberg 1-4, 6, NLP with FFNNs, DANs</td>
</tr>
<tr>
<td>Neural Nets II: NN impl / word</td>
<td>Goldberg 5, word2vec, GloVe, Dropout</td>
</tr>
<tr>
<td>embeddings</td>
<td></td>
</tr>
<tr>
<td>Neural Nets III: RNN and CNN</td>
<td>Goldberg 9-11, Kim</td>
</tr>
<tr>
<td>encoders</td>
<td></td>
</tr>
<tr>
<td>Neural Nets IV: Neural CRFs</td>
<td>Collobert and Weston, Neural NER, Neural CRF parsing</td>
</tr>
<tr>
<td>Trees II: Dependency I</td>
<td>JM 14.1-14.4, Huang 1-2</td>
</tr>
<tr>
<td>Trees III: Dependency II</td>
<td>Parsey, Huang 2</td>
</tr>
<tr>
<td>Semantics I</td>
<td></td>
</tr>
<tr>
<td>Semantics II / Seq2seq I</td>
<td></td>
</tr>
<tr>
<td>Seq2seq II: Beam search, attention</td>
<td>Seq2seq, Attention, Luong Attention</td>
</tr>
<tr>
<td>Information Extraction / SRL</td>
<td>Distant supervision, RL for slot filling, TextRunner, ReVerb, NELL</td>
</tr>
<tr>
<td>Discourse and Coreference</td>
<td></td>
</tr>
<tr>
<td>Machine Translation I: Phrase-based</td>
<td>HMM alignment, Pharaoh</td>
</tr>
<tr>
<td>Machine Translation II: Neural</td>
<td></td>
</tr>
<tr>
<td>Applications I: Reading</td>
<td>E2E Memory Networks, CBT, SQuAD, BiDAF</td>
</tr>
<tr>
<td>comprehension / MemNets</td>
<td></td>
</tr>
<tr>
<td>Applications II: Language</td>
<td></td>
</tr>
<tr>
<td>grounding</td>
<td></td>
</tr>
<tr>
<td>Applications III: Summarization</td>
<td>MMR, Gillick, Sentence compression, SummaRUNNER, Pointer</td>
</tr>
<tr>
<td>Applications IV: Dialogue</td>
<td>RNN chatbots, Diversity, Goal-oriented, Latent Intention, QA-as-dialogue</td>
</tr>
<tr>
<td>Unsupervised Learning</td>
<td></td>
</tr>
<tr>
<td>NO CLASS (Thanksgiving)</td>
<td></td>
</tr>
<tr>
<td>Multilinguality and morphology</td>
<td></td>
</tr>
<tr>
<td>Wrapup</td>
<td></td>
</tr>
</tbody>
</table>
### Outline of the Course

<table>
<thead>
<tr>
<th>Topics</th>
<th>Readings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction [40p]</td>
<td>JM 6.1-6.3</td>
</tr>
<tr>
<td>Binary classification</td>
<td>JM 7, Structured SVM secs 1-2</td>
</tr>
<tr>
<td>Multiclass classification</td>
<td>JM 9, JM 10.4, Manning POS</td>
</tr>
<tr>
<td>Sequence models I: HMMs</td>
<td>Sutton CRFs 2.3, 2.6.1, Wallach CRFs tutorial, Illinois NER</td>
</tr>
<tr>
<td>Sequence models II: CRFs</td>
<td>Goldberg 1-4, 6, NLP with FFNNs, DANs</td>
</tr>
<tr>
<td>Neural Nets I: FFNs</td>
<td>Goldberg 5, word2vec, GloVe, Dropout</td>
</tr>
<tr>
<td>Neural Nets II: NN impl / word embeddings</td>
<td>Goldberg 9-11, Kim</td>
</tr>
<tr>
<td>Neural Nets III: RNN and CNN encoders</td>
<td>Collobert and Weston, Neural NER, Neural CRF parsing</td>
</tr>
<tr>
<td>Neural Nets IV: Neural CRFs</td>
<td></td>
</tr>
<tr>
<td>Trees II: Dependency I</td>
<td>JM 14.1-14.4, Huang 1-2</td>
</tr>
<tr>
<td>Trees III: Dependency II</td>
<td>Parsey, Huang 2</td>
</tr>
<tr>
<td>Semantics I</td>
<td></td>
</tr>
<tr>
<td>Semantics II / Seq2seq I</td>
<td>Seq2seq, Attention, Luong Attention</td>
</tr>
<tr>
<td>Seq2seq II: Beam search, attention</td>
<td></td>
</tr>
<tr>
<td>Information Extraction / SRL</td>
<td>Distant supervision, RL for slot filling, TextRunner, ReVerb, NELL</td>
</tr>
<tr>
<td>Discourse and Coreference</td>
<td></td>
</tr>
<tr>
<td>Machine Translation I: Phrase-based</td>
<td>HMM alignment, Pharaoh</td>
</tr>
<tr>
<td>Machine Translation II: Neural</td>
<td></td>
</tr>
<tr>
<td>Applications I: Reading comprehension / MemNets</td>
<td>E2E Memory Networks, CBT, SQuAD, BiDAF</td>
</tr>
<tr>
<td>Applications II: Language grounding</td>
<td></td>
</tr>
<tr>
<td>Applications III: Summarization</td>
<td>MMR, Gillick, Sentence compression, SummaRUNNER, Pointer</td>
</tr>
<tr>
<td>Applications IV: Dialogue</td>
<td>RNN chatbots, Diversity, Goal-oriented, Latent Intention, QA-as-dialogue</td>
</tr>
<tr>
<td>Unsupervised Learning</td>
<td></td>
</tr>
<tr>
<td>NO CLASS (Thanksgiving)</td>
<td></td>
</tr>
<tr>
<td>Multilinguality and morphology</td>
<td></td>
</tr>
<tr>
<td>Wrapup</td>
<td></td>
</tr>
</tbody>
</table>
Outline of the Course

<table>
<thead>
<tr>
<th>Topics</th>
<th>Readings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction [40p]</td>
<td></td>
</tr>
<tr>
<td>Binary classification</td>
<td>JM 6.1-6.3</td>
</tr>
<tr>
<td>Multiclass classification</td>
<td>JM 7, Structured SVM secs 1-2</td>
</tr>
<tr>
<td>Sequence models I: HMMs</td>
<td>JM 9, JM 10.4, Manning POS</td>
</tr>
<tr>
<td>Sequence models II: CRFs</td>
<td>Sutton CRFs 2.3, 2.6.1, Wallach CRFs tutorial, Illinois NER</td>
</tr>
<tr>
<td>Neural Nets I: FFNNs</td>
<td>Goldberg 1-4, 6, NLP with FFNNs, DANs</td>
</tr>
<tr>
<td>Neural Nets II: NNP / word</td>
<td></td>
</tr>
<tr>
<td>embeddings</td>
<td>Goldberg 5, word2vec, GloVe, Dropout</td>
</tr>
<tr>
<td>Neural Nets III: RNN and CNN encoders</td>
<td>Goldberg 9-11, Kim</td>
</tr>
<tr>
<td>Neural Nets IV: Neural CRFs</td>
<td>Collorob and Weston, Neural NER, Neural CRF</td>
</tr>
<tr>
<td>Trees II: Dependency I</td>
<td>JM 14.1-14.4, Huang 1-2</td>
</tr>
<tr>
<td>Trees III: Dependency II</td>
<td>Parsey, Huang 2</td>
</tr>
<tr>
<td>Semantics I</td>
<td></td>
</tr>
<tr>
<td>Semantics II / Seq2seq I</td>
<td></td>
</tr>
<tr>
<td>Seq2seq II: Beam search, attention</td>
<td>Seq2seq, Attention, Luong Attention</td>
</tr>
<tr>
<td>Information Extraction / SRL</td>
<td>distant supervision, RL for slot filling, TextRunner, ReVerb, NELL</td>
</tr>
<tr>
<td>Discourse and Coreference</td>
<td></td>
</tr>
<tr>
<td>Machine Translation I: Phrase-based</td>
<td>HMM alignment, Pharaoh</td>
</tr>
<tr>
<td>Machine Translation II: Neural</td>
<td></td>
</tr>
<tr>
<td>Applications I: Reading comprehension / MemNets</td>
<td>E2E Memory Networks, CBT, SQuAD, BiDAF</td>
</tr>
<tr>
<td>Applications II: Language grounding</td>
<td></td>
</tr>
<tr>
<td>Applications III: Summarization</td>
<td>MMR, Gillick, Sentence compression, SummaRunNER, Pointer</td>
</tr>
<tr>
<td>Applications IV: Dialogue</td>
<td>RNN chatbots, Diversity, Goal-oriented, Latent Intention, QA-as-dialogue</td>
</tr>
<tr>
<td>Unsupervised Learning</td>
<td></td>
</tr>
<tr>
<td>NO CLASS (Thanksgiving)</td>
<td></td>
</tr>
<tr>
<td>Multilinguality and morphology</td>
<td></td>
</tr>
<tr>
<td>Wrapup</td>
<td></td>
</tr>
</tbody>
</table>
# Outline of the Course

## ML and structured prediction for NLP

<table>
<thead>
<tr>
<th>Topics</th>
<th>Readings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction [4pp]</td>
<td></td>
</tr>
<tr>
<td>Binary classification</td>
<td>JM 6.1-6.3</td>
</tr>
<tr>
<td>Multiclass classification</td>
<td>JM 7, Structured SVM secs 1-2</td>
</tr>
<tr>
<td>Sequence models I: HMMs</td>
<td>JM 9, JM 10.4, Manning POS</td>
</tr>
<tr>
<td>Sequence models II: CRFs</td>
<td>Sutton CRFs 2.3, 2.6.1, Wallach CRFs tutorial, Illinois NER</td>
</tr>
<tr>
<td>Neural Nets I: FFNNs</td>
<td>Goldberg 1-4, 6, NLP with FFNNs, DANs</td>
</tr>
<tr>
<td>Neural Nets II: NN impl/word embeddings</td>
<td>Goldberg 5, word2vec, GloVe, Dropout</td>
</tr>
<tr>
<td>Neural Nets III: RNN and CNN encoders</td>
<td>Goldberg 9-11, Kim</td>
</tr>
<tr>
<td>Neural Nets IV: Neural CRFs</td>
<td>Collobot and Weston, Neural NER, Neural CRF parsing</td>
</tr>
<tr>
<td>Trees II: Dependency I</td>
<td>JM 14.1-14.4, Huang 1-2</td>
</tr>
<tr>
<td>Trees III: Dependency II</td>
<td>Parsey, Huang 2</td>
</tr>
<tr>
<td>Semantics I</td>
<td></td>
</tr>
<tr>
<td>Semantics II/Seq2seq I</td>
<td></td>
</tr>
<tr>
<td>Seq2seq II: Beam search, attention</td>
<td>Seq2seq, Attention, Luong Attention</td>
</tr>
<tr>
<td>Information Extraction/SRL</td>
<td>Distant supervision, RL for slot filling, TextRunner, ReVerb, NELL</td>
</tr>
<tr>
<td>Discourse and Coreference</td>
<td></td>
</tr>
<tr>
<td>Machine Translation I: Phrase-based</td>
<td>HMM alignment, Pharaoh</td>
</tr>
<tr>
<td>Machine Translation II: Neural</td>
<td></td>
</tr>
<tr>
<td>Applications I: Reading comprehension/MemNets</td>
<td>E2E Memory Networks, CBT, SQuAD, BiDAF</td>
</tr>
<tr>
<td>Applications II: Language grounding</td>
<td></td>
</tr>
<tr>
<td>Applications III: Summarization</td>
<td>MMR, Gillick, Sentence compression, SummaRUNNER, Pointer</td>
</tr>
<tr>
<td>Applications IV: Dialogue</td>
<td>RNN chatbots, Diversity, Goal-oriented, Latent Intention, QA-as-dialogue</td>
</tr>
<tr>
<td>Unsupervised Learning</td>
<td></td>
</tr>
<tr>
<td>NO CLASS (Thanksgiving)</td>
<td></td>
</tr>
<tr>
<td>Multilinguality and morphology</td>
<td></td>
</tr>
<tr>
<td>Wrapup</td>
<td></td>
</tr>
</tbody>
</table>
Course Goals
Course Goals

- Cover fundamental machine learning techniques used in NLP
Course Goals

- Cover fundamental machine learning techniques used in NLP
- Understand how to look at language data and approach linguistic phenomena
Course Goals

- Cover fundamental machine learning techniques used in NLP
- Understand how to look at language data and approach linguistic phenomena
- Cover modern NLP problems encountered in the literature: what are the active research topics in 2018?
Course Goals

- Cover fundamental machine learning techniques used in NLP
- Understand how to look at language data and approach linguistic phenomena
- Cover modern NLP problems encountered in the literature: what are the active research topics in 2018?
- Make you a “producer” rather than a “consumer” of NLP tools
Course Goals

- Cover fundamental machine learning techniques used in NLP
- Understand how to look at language data and approach linguistic phenomena
- Cover modern NLP problems encountered in the literature: what are the active research topics in 2018?
- Make you a “producer” rather than a “consumer” of NLP tools
  - The four assignments should teach you what you need to know to understand nearly any system in the literature
Assignments

- 4 Homework Assignments
  - Implementation-oriented, with an open-ended component to each
  - Homework 1 (Naive Bayes for sentiment classification) is out NOW
  - ~2 weeks per assignment, 3 “slip days” for automatic extensions
Assignments

- 4 Homework Assignments
  - Implementation-oriented, with an open-ended component to each
  - Homework 1 (Naive Bayes for sentiment classification) is out NOW
  - ~2 weeks per assignment, 3 “slip days” for automatic extensions

These projects require understanding of the concepts, ability to write performant code, and ability to think about how to debug complex systems. They are challenging, so start early!
Final Project

- Final project (20%)
  - Groups of 3-4 preferred, 1 is possible.
  - Good idea to talk to run your project idea by me in office hours or email.
  - 4 page report + final project presentation.