CSE P 517
Natural Language Processing
Winter 2021

Introduction
Yejin Choi

Slides adapted from Dan Klein, Luke Zettlemoyer
What is NLP?

Fundamental goal: *deep* understand of *broad* language
- Not just string processing or keyword matching

End systems that we want to build:
- Simple: spelling correction, text categorization…
- Complex: speech recognition, machine translation, information extraction, sentiment analysis, question answering…
- Unknown: human-level comprehension (is this just NLP?)
Some examples of the language groups

<table>
<thead>
<tr>
<th>Language Group</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Afro-Asiatic</td>
<td>Khoisan, Khoisan languages, Khoisanic languages</td>
</tr>
<tr>
<td>Niger-Congo</td>
<td>Bantu, Nilo-Saharan, Khoisan, Indo-European</td>
</tr>
<tr>
<td>Altaic</td>
<td>Turkic, Mongolic, East Siberian languages</td>
</tr>
<tr>
<td>Uralic</td>
<td>Finno-Ugric, Ugric, Finnic, Uralic</td>
</tr>
<tr>
<td>Dravidian</td>
<td>Dravidian, Dravidan languages, Dravidanic</td>
</tr>
<tr>
<td>Sino-Tibetan</td>
<td>Chinese, Burmese-Tibetan, Tibetan</td>
</tr>
<tr>
<td>Austro-Asiatic</td>
<td>Austronesian, Borneo-Philippines/Formosan</td>
</tr>
<tr>
<td>Na-Dene</td>
<td>Na-Dene, Athabaskan, Algonquian</td>
</tr>
<tr>
<td>Eskimo-Aleut</td>
<td>Eskimo, Aleut, Alutue, Alutian</td>
</tr>
<tr>
<td>American Indian</td>
<td>Algonquian, Iroquoian, Washo</td>
</tr>
<tr>
<td>Aic</td>
<td>Aic, Austronesian</td>
</tr>
<tr>
<td>Uto-Aztecan</td>
<td>Uto-Aztecan, Athabaskan</td>
</tr>
<tr>
<td>Andean</td>
<td>Andean, Quechua</td>
</tr>
<tr>
<td>Tupian</td>
<td>Tupian, Mbuti</td>
</tr>
<tr>
<td>Brazilian indigenous</td>
<td>Brazilian, Guarani</td>
</tr>
<tr>
<td>Isolate</td>
<td>Isolate, isolate languages</td>
</tr>
</tbody>
</table>
Machine Translation

- Translate text from one language to another
- Recombines fragments of example translations
- Challenges:
  - What fragments? [learning to translate]
  - How to make efficient? [fast translation search]
  - Fluency (second half of this class) vs fidelity (later)
Impôt sur le revenu : vous en 2014 ?

Sélectionnez votre revenu et votre situation pour bénéficier de la pause fiscale.
- Comment le budget pour 2014 est-il réparti ?
- Un budget 2014 soumis aux critiques

Income tax: how much do you pay in 2014?

Select your income and family situation to see if you get the tax break.
- How is the budget for 2014 allocated?
- A 2014 budget submitted to criticism
- Budget: these expenses no government can reduce
- Budget 2014: the retail savings

Unemployment fell for the first time since April 2011
Surviving in the Central time looting and anarchy

DÉCOUVREZ TOUS LES SERVICES ABONNÉS
S'abonner au Monde à partir de 1 €

Member(s) of Europe Ecology-Greens, do you share the finding of severe Christmas Mamère EELV?
Share your experience

Continuous
7:53 Budget: the fixed expenses
7:36 Heard the "Fashion Week" in Paris
7:19 Control giant Airbus
7:04 Complaint against "Actual Values"
7:01 Venezuela: 17 people arrested
6:59 Vidberg: the new budget came
6:50 The "noble mission" of the NSA
6:38 Roma: jousting between Brussels...

Patchwork of strategies, Kenia, Tunisia, India, Pakistan, Ukraine, China, Ethiopia

International Labor Day - 1er Mai 2013

Le chômage baisse pour la première fois depuis avril 2011

POST DE BLOG

Surviving in the Central time looting and anarchy

DE FURSAC

automne-hiver 13/14
"Обиженные люди работают, а иностранцы нам не поедут"
25.09.2013 19:48

"Mentally ill people are working, and foreign scholars to us will not go"
25.09.2013 19:48

"against all"
05/29/2013 13:37

World through the lens: September 25.
25/09/2013 10:27
US Cities: Its largest airport is named for a World War II hero; its second largest, for a World War II battle.
Knowledge Graph: “things not strings”

The Knowledge Graph
Learn more about one of the key breakthroughs behind the future of search.

See it in action
Discover answers to questions you never thought to ask, and explore collections and lists.
New York Times Co. named Russell T. Lewis, 45, president and general manager of its flagship New York Times newspaper, responsible for all business-side activities. He was executive vice president and deputy general manager. He succeeds Lance R. Primis, who in September was named president and chief operating officer of the parent.

<table>
<thead>
<tr>
<th>Person</th>
<th>Company</th>
<th>Post</th>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td>Russell T. Lewis</td>
<td>New York Times newspaper</td>
<td>president and general manager</td>
<td>started</td>
</tr>
<tr>
<td>Russell T. Lewis</td>
<td>New York Times newspaper</td>
<td>executive vice president</td>
<td>ended</td>
</tr>
<tr>
<td>Lance R. Primis</td>
<td>New York Times Co.</td>
<td>president and CEO</td>
<td>started</td>
</tr>
</tbody>
</table>
Information Extraction

New York Times Co. named Russell T. Lewis, 45, president and general manager of its flagship New York Times newspaper, responsible for all business-side activities. He was executive vice president and deputy general manager. He succeeds Lance R. Primis, who in September was named president and chief operating officer of the parent.

<table>
<thead>
<tr>
<th>Person</th>
<th>Company</th>
<th>Post</th>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td>Russell T. Lewis</td>
<td>New York Times newspaper</td>
<td>president and general manager</td>
<td>started</td>
</tr>
<tr>
<td>Russell T. Lewis</td>
<td>New York Times newspaper</td>
<td>executive vice president</td>
<td>ended</td>
</tr>
<tr>
<td>Lance R. Primis</td>
<td>New York Times Co.</td>
<td>president and CEO</td>
<td>started</td>
</tr>
</tbody>
</table>

Sub-problems:

1) Named entity recognition: finding named entities X and their types T(X)
   persons: “Russell T. Lewis”, “Lance R. Primis”

2) Relation extraction: the relation R(X,Y) between named entities X, Y
   Works_for(Russell T. Lewis, New York Times Newspaper)

3) Coreference resolution: which text spans refer to the same named entity?
   {Russell T.Lewis, He, He} are an equivalence set.

- Is this easy or hard?
- Easier if the model exploits the redundancy of information!
Question Answering:
- More than search
- Can be really easy: “What’s the capital of Wyoming?”
- Can be harder: “How many US states’ capitals are also their largest cities?”
- Can be open ended: “What are the main issues in the global warming debate?”

Natural Language Interaction:
- Understand requests and act on them
- “Make me a reservation for two at Quinn’s tonight”

capital of Wyoming: Information From Answers.com
Note: click on a word meaning below to see its connections and related words
The noun capital of Wyoming has one meaning: Meaning #1: the capital.
www.answers.com/topic/capital-of-wyoming - 21k - Cached - Similar pages

Cheyenne: Weather and Much More From Answers.com
Cheyenne (ˈʃi-ən) The capital of Wyoming, in the southeast part of the state near the Nebraska and Colorado borders.
www.answers.com/topic/cheyenne-wyoming - 74k - Cached - Similar pages
Human-Machine Interactions

“What’s the best movie to see this weekend”

That would probably start an argument. But here’s a list of highly-regarded movies:

25 MOVIES

NORTH BY NORTHWEST
Released July 17, 1959
100%

THE TREASURE OF THE SIERRA...
Released January 6, 1948
100%

What can I help you with?

“You need to start understanding me Siri”

I’ll make a note of that.

“Yeah you better make a note of that”

Noted:

Of that
Will this Be Part of All Our Home Devices?

- Will it rain tomorrow?
- Set an alarm for eight a.m.
- Play music by Bruno Mars
- How many teaspoons are in a tablespoon?
- Add gelato to my shopping list
- Wikipedia: Abraham Lincoln
- When is Thanksgiving?
- Play my “dinner party” playlist
- What’s the weather in Los Angeles this weekend?
- Add “make hotel reservations” to my to-do list
The Alexa Prize
$2.5 Million to Advance Conversational Artificial Intelligence
September 2016 – November 2017

Amazon Selects Teams to Compete for Inaugural $2.5 Million Alexa Prize
• UW Sounding Board among 3 Finalists!
• Final competition in Las Vegas in Nov
• Unclear if any team will make the 20 min goal
• How not to win:
  – Brute force more data, more depth
  – Add RL and pray magic will arise
Announced at AWS re:INVENT
Speech Recognition

- **Automatic Speech Recognition (ASR)**
  - Audio in, text out
  - SOTA: 0.3% error for digit strings, 5% dictation, 50%+ TV

- **Text to Speech (TTS)**
  - Text in, audio out
  - SOTA: totally intelligible (if sometimes unnatural)

"Speech Lab"
Analyzing public opinion, making political forecasts

- Today: In 2012 election, automatic sentiment analysis actually being used to complement traditional methods (surveys, focus groups)
- Past: “Sentiment Analysis” research started in 2002
- Future: computational social science and NLP for digital humanities (psychology, communication, literature and more)
- Challenge: Need statistical models for deeper semantic understanding --- subtext, intent, nuanced messages
Summarization

- Condensing documents
  - Single or multiple docs
  - Extractive or synthetic
  - Aggregative or representative

- Very context-dependent!

- An example of analysis with generation

WASHINGTON (CNN) -- President Obama's inaugural address was cooler, more measured and reassuring than that of other presidents making it, perhaps, the right speech for the times. Some inaugural addresses are known for their soaring, inspirational language. Like John F. Kennedy's in 1961: "Ask not what your country can do for you. Ask what you can do for your country."

Obama's address was less stirring, perhaps, but it was also more candid and down-to-earth. "Starting today, the new president said. "We must begin to make our way back."

"We gather because we have chosen hope over fear, unity of purpose over conflict and discord," Obama said.

Obama's call to unity after decades of political division echoed Abraham Lincoln's first inaugural address in 1861. Even though he delivered it at the onset of a terrible civil war, Lincoln's speech was not a call to battle. It was a call to look beyond the war, toward reconciliation based on what he called "the better angels of our nature."

Some presidents used their inaugural address to set out a bold agenda.
Some of the formulaic news articles are now written by computers.

- Definitely far from "Op-ed"
- Can we make the generation engine statistically learned rather than engineered?
Despite an expected dip in profit, analysts are generally optimistic about **Steelcase** as it prepares to report its third-quarter earnings on Monday, December 22, 2014. The consensus earnings per share estimate is 26 cents per share.

The consensus estimate remains unchanged over the past month, but it has decreased from three months ago when it was 27 cents. Analysts are expecting earnings of 85 cents per share for the fiscal year. Revenue is projected to be 5% above the year-earlier total of $784.8 million at $826.1 million for the quarter. For the year, revenue is projected to come in at $3.11 billion.

The company has seen revenue grow for three quarters straight. The less than a percent revenue increase brought the figure up to $786.7 million in the most recent quarter. Looking back further, revenue increased 8% in the first quarter from the year earlier and 8% in the fourth quarter.

The majority of analysts (100%) rate Steelcase as a buy. This compares favorably to the analyst ratings of three similar companies, which average 57% buys. Both analysts rate Steelcase as a buy.

Steelcase is a designer, marketer and manufacturer of office furniture. Other companies in the furniture and fixtures industry with upcoming earnings release dates include: HNI and Knoll.
“Imagine, for example, a computer that could look at an arbitrary scene anything from a sunset over a fishing village to Grand Central Station at rush hour and produce a verbal description. This is a problem of overwhelming difficulty, relying as it does on finding solutions to both vision and language and then integrating them. I suspect that scene analysis will be one of the last cognitive tasks to be performed well by computers”

-- David Stork (HAL’s Legacy, 2001) on A. Rosenfeld’s vision
What begins to work (e.g., Kuznetsova et al. 2014)

The flower was so vivid and attractive.

Blue flowers are running rampant in my garden.

Spring in a white dress.

Blue flowers have no scent. Small white flowers have no idea what they are.

Scenes around the lake on my bike ride.

This horse walking along the road as we drove by.

We sometimes do well: 1 out of 4 times, machine captions were preferred over the original Flickr captions:
But many challenges remain (better examples of when things go awry)

The couch is definitely bigger than it looks in this photo.

My cat laying in my duffel bag.

Yellow ball suspended in water.

A high chair in the trees.
Table of Content

- Definition of NLP
- Historical account of NLP
NLP History: pre-statistics

(1) Colorless green ideas sleep furiously.
(2) Furiously sleep ideas green colorless.

- It is fair to assume that neither sentence (1) nor (2) (nor indeed any part of these sentences) had ever occurred in an English discourse. Hence, in any statistical model for grammaticalness, these sentences will be ruled out on identical grounds as equally "remote" from English. Yet (1), though nonsensical, is grammatical, while (2) is not.” (Chomsky 1957)

- 70s and 80s: more linguistic focus
  - Emphasis on deeper models, syntax and semantics
  - Toy domains / manually engineered systems
  - Weak empirical evaluation
NLP: machine learning and empiricism

“Whenever I fire a linguist our system performance improves.” –Jelinek, 1988

- **1990s: Empirical Revolution**
  - Corpus-based methods produce the first widely used tools
  - Deep linguistic analysis often traded for robust approximations
  - *Empirical evaluation* is essential

- **2000s: Richer linguistic representations used in statistical approaches, scale to more data!**

- **2010s: you decide!**
What is Nearby NLP?

- **Computational Linguistics**
  - Using computational methods to learn more about how language works
  - We end up doing this and using it

- **Cognitive Science**
  - Figuring out how the human brain works
  - Includes the bits that do language
  - Humans: the only working NLP prototype!

- **Speech?**
  - Mapping audio signals to text
  - Traditionally separate from NLP, converging?
  - Two components: acoustic models and language models
  - Language models in the domain of stat NLP
Table of Content

- Definition of NLP
- Historical account of NLP
- Unique challenges of NLP
Problem: Ambiguities

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

- Why are these funny?
Hurricane Emily howled toward Mexico's Caribbean coast on Sunday packing 135 mph winds and torrential rain and causing panic in Cancun, where frightened tourists squeezed into musty shelters.

- **SOTA:** ~90% accurate for many languages when given many training examples, some progress in analyzing languages given few or no examples.
Semantic Ambiguity

At last, a computer that understands you like your mother.

- **Direct Meanings:**
  - It understands you like your mother (does) [presumably well]
  - It understands (that) you like your mother
  - It understands you like (it understands) your mother

- **But there are other possibilities, e.g. mother could mean:**
  - a woman who has given birth to a child
  - a stringy slimy substance consisting of yeast cells and bacteria; is added to cider or wine to produce vinegar

- **Context matters, e.g. what if previous sentence was:**
  - Wow, Amazon predicted that you would need to order a big batch of new vinegar brewing ingredients. 😊

[Example from L. Lee]
Dark Ambiguities

- **Dark ambiguities**: most structurally permitted analyses are so bad that you can’t get your mind to produce them

This analysis corresponds to the correct parse of

“This will panic buyers!”

- **Unknown words and new usages**
- **Solution**: We need mechanisms to focus attention on the best ones, probabilistic techniques do this
Problem: Scale

- People *did* know that language was ambiguous!
  - ...but they hoped that all interpretations would be “good” ones (or ruled out pragmatically)
  - ...they didn’t realize how bad it would be
Corpora

- A corpus is a collection of text
  - Often annotated in some way
  - Sometimes just lots of text
  - Balanced vs. uniform corpora

- Examples
  - Newswire collections: 500M+ words
  - Brown corpus: 1M words of tagged “balanced” text
  - Penn Treebank: 1M words of parsed WSJ
  - Canadian Hansards: 10M+ words of aligned French / English sentences
  - The Web: billions of words of who knows what
Problem: Sparsity

- However: sparsity is always a problem
  - New unigram (word), bigram (word pair)
Table of Content

- Definition of NLP
- Historical account of NLP
- Unique challenges of NLP
- Class administrivia
Site & Crew

- Site: [https://courses.cs.washington.edu/courses/csep517/21wi/](https://courses.cs.washington.edu/courses/csep517/21wi/)
- Canvas: [https://canvas.uw.edu/courses/1445050](https://canvas.uw.edu/courses/1445050)
- Crew:
- Instructor: [Yejin Choi](mailto:Yejin.Choi@uw.edu) (office hour: Mon/Wed 8:00 – 8:30pm)
- TA:
  - Xiujun Li
  - James Park
  - Peter West
  - Ximing Lu
Textbooks and Notes

- **Textbook (recommended but not required):**

- Lecture slides & notes are required
  - See the course website for details

- **Assumed Technical Background:**
  - Data structure, algorithms, strong programming skills, probabilities, statistics
Grading & Policy

Grading:
- 5 programming assignments (75%)
- Self-defined final project (15%)
- Participation (in-class and online forums) (10%)

Policy:
- All homework will be completed individually.
- Final projects can be done in groups of 1-3 people.
- Plagiarism will lead to grave consequences.

Participation and Discussion:
- Class participation is expected and appreciated!!!
- Email is great, but please use the message board when possible (we monitor it closely)
What is this Class?

- Three aspects to the course:
  - Linguistic Issues
    - What are the range of language phenomena?
    - What are the knowledge sources that let us disambiguate?
    - What representations are appropriate?
    - How do you know what to model and what not to model?
  - Statistical Modeling Methods
    - Increasingly complex model structures
    - Learning and parameter estimation
    - Efficient inference: dynamic programming, search, sampling
  - Engineering Methods
    - Issues of scale
    - Where the theory breaks down (and what to do about it)
- We’ll focus on what makes the problems hard, and what works in practice…
# Approximate Schedule

<table>
<thead>
<tr>
<th>Week</th>
<th>Dates</th>
<th>Topics &amp; Lecture Slides</th>
</tr>
</thead>
</table>
| 1    | Jan 4, 6         | **Introduction** [slides (old)]
|      |                  | **Probabilistic Models of Words:** N-gram Language Models (LMs) [slides (old)] |
| 2    | Jan 11, 13       | **Probabilistic Models of Sequences:** Hidden Markov Models (HMMs) [slides (old)] |
| 3    | Jan 20 (Jan 18: MLK day) | **Probabilistic Models of Sequences:** Hidden Markov Models (HMMs) & EM [slides (old)] |
| 4    | Jan 25, 27       | **Learning (Feature-Rich Models):** Log-Linear Models, Maximum Entropy Models [slides (old)] |
| 5    | Feb 1, 3         | **Learning (Structural Graphical Models):** Conditional Random Fields (CRFs) [slides (old)] |
| 6    | Feb 8, 10        | **Neural Language Models:** Basics [slides (old)] |
| 7    | Feb 17 (Feb 15: Presidents Day) | **Neural Language Models:** Transformers |
| 8    | Feb 22, 24       | **Neural Language Generation:** Summarization, Dialogues, and De-generation |
| 9    | Mar 1, 3         | **Neural Knowledge and Reasoning:** On Commonsense Intelligence |
| 10   | Mar 8, 10        | **New Challenges with Neural Models:** Interpretability, Bias, and Ethics |
Comparisons with Other Classes

- Compared to ML
  - Typically multivariate, dynamic programming everywhere
  - Structural Learning & Inference
  - Insights into language matters (a lot!)
  - DL: RNNs, LSTMs, Seq-to-seq, Attention, …

- Compared to CompLing classes
  - More focus on core algorithm design, technically more demanding in terms of math, algorithms, and programming
Class Requirements and Goals

- **Class requirements**
  - Uses a variety of skills / knowledge:
    - Probability and statistics
    - Decent coding skills
    - Data structure and algorithms (dynamic programming!)
    - (Optional) basic linguistics background
  - ML/AI helps if you’ve taken either before, but not necessary

- **Class goals**
  - Learn the fundamental concepts and techniques
  - Learn current engineering practices
  - Learn how to advance the field!