## CSE 517 Natural Language Processing - Winter 2018! -

Yejin Choi Computer Science & Engineering WUNIVERSITY of WASHINGTON

### What is NLP like today?



Do we know how to teach language?

Yes! for humans; Not so well for machines

#### Which of these is the hardest for humans?

- summarizing a children's book in a few sentences
- 2. making a small talk with a child
- 3. reading a movie script and answering a question about the story
- 4. reading a wikipedia article and answering a question about the article
- 5. translating a Korean text to a Polish text

#### Which of these is the hardest for machines?

- summarizing a children's book in a few sentences
- 2. making a small talk with a child
- 3. reading a movie script and answering a question about the story
- 4. reading a wikipedia article and answering a question about the article
- 5. translating a Korean text to a Polish text

### Machine Translation



- How to automatically induce the word-level or phraselevel alignments between two languages?
- (without learning how to understand either language properly)

# Machine Translation (2013 google translate)

Impôts Kenya Syrie EN CE MOMENT

Taxes Kenya Syria Pakistan Use Prism scandal AT THIS MOMENT

#### Impôt sur le revenu vous en 2014?



Sélectionnez votre revenu et votre situ bénéficiez de la pause fiscale.

- Comment le budget pour 2014 est-il réparti ? .. visuel interactif
- Un budget 2014 soumis aux critiques



Le chômage baisse pour la première fois depuis avril 2011 POST DE BLOG

#### 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 is allocated? .il INTERACTIVE VISUAL
- A 2014 budget submitted to criticism
- Budget: these expenses no government can reduce
- Budget 2014: the retail savings ... **INTERACTIVE VISUAL**



#### **CALL FOR EVIDENCE**

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

Share your experience

#### Continuous

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



automne-hiver 13/14



Unemployment fell for the first time since April 2011 POST BLOG



Surviving in the Central time

looting and anarchy

## Speech Translation

- Automatic translation
  - -- not perfect, but good enough for people to use
  - -- real time translation with audio
  - -- first statistical model (IBM model 1) came out in 1993
  - -- first MT service based on statistical model in 2007



## Information Search & Extraction

- Web search today can handle natural language queries better
- often presents us structured knowledge

	Everything	Vincent van Gogh - Wikipedia, the free encyclopedia en.wikipedia.org/wiki/Vincent_van_Gogh Vincent Willem van Gogh was a Dutch post-Impressionist painter whose work, notable for its rough beauty, emotional honesty, and bold color, had a far-reaching :-> List of works by Vincent van The Starry Night - Vincent van Gogh's health Vincent van Gogh Gallery - Welcome! www.vangoghgallery.com/ The definitive reference for information about Vincent van Gogh including his biography and the complete collection of his paintings, drawings, sketches and	Vincent van Gogh				
	Images			Vincent Willem van Gogh was a Dutch post-			
	Maps		rough beauty, emotional honesty, and bold color, had a far-reaching influence on 20th-century art. Wikipedia				
	Videos						
	News						
				Born: March 30, 1853, Zundert			
	Shopping		Died: July 29, 1890, Auvers-sur-Oise				
	More			Parents: Anna Carbentus van Gogh			
		Vincent van Gogh Biography - His Life and Times		Siblings: 1	Theo van G	logh, Wil van G	ogh
(	Portland, ME	www.vangoghgallery.com/misc/bio.html Read a biography of Dutch post-Impressionist artist Vincent van Gogh. Get quick facts, a timeline, information about his family and artists who influenced him.		Periods: Expressionism, Divisionism, Post- Impressionism, Impressionism			
	Change location						
			Works				
	All results	Images for vincent van gogh - Report images	AND DO NOT		the second second	The state of the	10000
	Sites with images				J. 15	1000	-
	More search tools			A. 11 N		Pro de	THU:
			The Stamy Cale	State Street Street Street	Detate	in an	Clamy Mail
			Night Terra	ce at Eat	ers	1889	Over the

WebMuseum: Gogh, Vincent van www.ibiblio.org/wm/paint/auth/gogh/ Provides information and images on some of his works.

#### The Vincent van Gogh Gallery

www.vggallery.com/

A comprehensive resource for information about Van Gogh and images of his works. Has images of all the paintings, sketches, watercolours, letter sketches, and ...

Vincent van Gogh (1853–1890) | Thematic Essay | Heilbrunn ... www.metmuseum.org/toah/hd/gogh/hd\_gogh.htm

Leonardo da

Vinci

1885

Paul

Gauguin

Night

Claude

Monet

1888

People also search for

1889

Picasso

Report a problem

1888

#### Knowledge Graph: "things not strings"

Home Tips & Tricks

**Features** Search Stories

Playground Blog Help



#### **Question Answering**



US Cities: Its largest airport is named for a World War II hero; its second largest, for a World War II battle.

#### Jeopardy! World Champion



#### Conversation with Devices

68%





#### $\blacksquare$

#### The Alexa Prize

#### \$2.5 Million to Advance Conversational Artificial Intelligence

September 2016 – November 2017



Conversational AI with long-term coherence

- Grand challenge: 20 minutes
- My initial guess: 1-2 minutes
- Our (winning) system --- 10+ minutes



#### system architecture? sorry, not this kind:





#### 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





# Language and Vision



"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



Rehard O. Dads Peter E. Hare David G. Stool Pattern Classification

#### What begins to work (e.g., Kuznetsova et al. 2014)











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.



#### But many challenges remain (better examples of when things go awry)



## How did NLP begin?

## 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's in the class?

#### Buffalo buffalo Buffalo buffalo buffalo buffalo Buffalo buffalo

#### WIKIW @RLD\* by Grag Orthiams

#### Buffalo buffalo Buffalo buffalo buffalo buffalo buffalo Buffalo buffalo.

is a grammatically correct sentence used as an example of how homonyms and homophones can be used to Create complicated constructs. The sentence is unpunctuated and uses three different readings of the word "buffalo." In order of their first use, these are:

- The city of Buffalo, New York.
- The animal "buffalo," in the plural (equivalent to "buffaloes"), in order to avoid articles.
- The verb "buffalo," meaning to confuse, deceive or intimidate.



Substituting the synonym "bison" for "buffalo" (animal), "bully" for "buffalo" (verb) and leaving "Buffalo" to mean the city, yields:

Buffalo bison, whom other Buffalo bison bully, themselves bully Buffalo bison. BUFFALD BIFFALD BIFFAL

distinct meanings

Homophone = a word

same as another word

but differs in meaning

which is pronounced the

00



Text excerpted from the Wikipedia articles Buffalo buffalo Buffalo buffalo buffalo buffalo buffalo., Homonym and Homophone. 26 March 2007

#### Probabilistic Models of Language

- Is it possible to model p(x), where x is a sentence of any length with any words such that p(x) is a valid probability distribution?
- Is it possible to automatically infer linguistic categories of words (part of speech) just by reading lots of text with no supervision?
- Is it possible to automatically infer linguistic structure of sentences just by reading lots of text with no supervision?

#### Neural network models of language

(Google NMT Oct 2016)



## 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?

## Syntactic Analysis



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)



#### Class Administrivia

### Site & Crew

- Site: <u>https://courses.cs.washington.edu/courses/cse517/19wi/</u>
- Canvas: <u>https://canvas.uw.edu/courses/1254676/</u>
- Crew:
- Instructor:

<u>Yejin Choi</u> (office hour: Thu 4:30 – 5:30) --- except this week: Thu 5:15 – 6:15

TA:

Hannah Rashkin Max Forbes Rowan Zellers







## Textbooks and Notes

- Textbook (recommended but not required):
  - Jurafsky and Martin, Speech and Language Processing, 2<sup>nd</sup> Edition
  - Manning and Schuetze, Foundations of Statistical NLP
  - GoodFellow, Bengio, and Courville, "Deep Learning" (free online book available at <u>deeplearningbook.org</u>)
- Lecture slides & notes are required
  - See the course website for details
- Assumed Technical Background:
  - Data structure, algorithms, strong programming skills, probabilities, statistics





## 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

1	I. Introduction				
	II. <b>Words</b> : Language Models (LMs)				
2	II. <b>Words</b> : Unknown Words (Smoothing)				
2	III. <b>Sequences</b> : Hidden Markov Models (HMMs)				
3	III. <b>Sequences</b> : Hidden Markov Models (HMMs) & EM				
Д	V. <b>Trees</b> : Probabilistic Context Free Grammars (PCFG)				
	V. <b>Trees</b> : Grammar Refinement				
Б	V. <b>Trees</b> : Dependency Grammars				
5	IV. Learning (Feature-Rich Models): Log-Linear Models				
6	IV. Learning (Structural Graphical Models): Conditional Random Fields (CRFs)				
7	VII. <b>Semantics</b> : Frame Semantics				
/	VII. Semantics: Distributed Semantics, Embeddings				
8	VIII. Deep Learning: Neural Networks				
9	VIII. <b>Deep Learning</b> : More NNs				
10	VIII. Deep Learning: Yet More NNs				

# Grading & Policy

- Grading:
  - 4 homework (55%)
  - In-class workbook (10%)
  - final project (30%)
  - course/discussion board participation (5%)

#### Policy:

- All homework will be completed individually.
- Final projects can be done in groups.
- Academic honest and plagiarism.
- Participation and Discussion:
  - Class participation is expected and appreciated!!!
  - Email is ok, but we prefer the message board at Canvas whenever possible





## Homework (55%)

Four major programming assignments:

- 1. Language Models (10%)
  - Conditional probabilities
  - Handling of unknown words & smoothing
- 2. HMMs (15%)
  - Viterbi algorithm with longer context
  - Forward backward & EM (bonus)
- 3. Structured Inference (15%)
  - How to convert a simple perceptron to structured perceptron
- 4. Deep Learning (15%)
  - Reading comprehension with pytorch

# Project (30%)

- Final project proposal (5%)
- Final project poster presentation (12%)
- Final project report (13%)
- Work as a team of 1 3 people
- Must contain some NLP components
- Ok to recycle your current research project

## 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!

## 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
- Compared to 447 / 547
  - 70% overlap depending on who taught the class

### Add Code / Audit

- Sorry, the class has been overbooked for a while
  - higher priorities on PhD students in ECE & linguistics
  - grads in other fields: please consider CompLing classes or CSE 447/547
  - ugrads in CSE: please take 447/547!
- Audit ok if there are seats (still) not taken