

Natural Language Processing

Introduction
Spring 2019

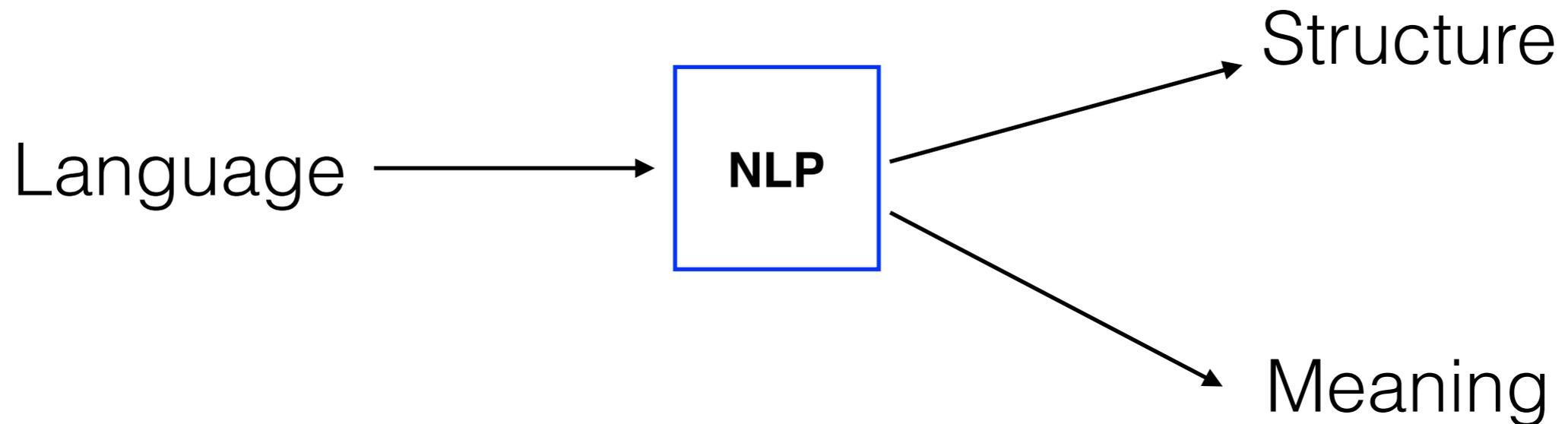
2001: A Space Odyssey

In case you didn't notice

- The class will be in English

What is NLP?

- **Goal:** Develop methods for processing, analyzing and understanding the **structure** and **meaning** of language.



- For all languages
- Many “levels”: transcribed, newswire ,social media...

What is NLP?

- **Application:** Build systems that help people do stuff (with text):
 - Question answering, virtual assistants, automatic translation, etc.
- There is a lot of language out there...

Applications are taking off!

- Search
- Translation
- Sentiment analysis
- Speech recognition
- Chatbots
- Virtual assistants
- Ad matching
- ...

Levels of analysis

- **Phonology**: sounds that make up language (more focus on the speech community)
 - *da'vid ben gur'jo:n: דוד בן-גוריון*
- **Morphology**: internal structure of words
 - *ובבתיהם*
- **Syntax**: structure of phrases, how words modify one another
 - *pretty little girl's school*
- **Semantics**: meaning of language in the world
 - *students on the back row*
- **Discourse**: relations between clauses and sentences
 - *I registered to the class on NLP **because** it is fascinating*
- **Pragmatics**: the way we use language in context
 - *It's hot in here*



What's so special about language

- Created by humans for communication

- מותר האדם מן הבהמה

- Learned from experience (!!)

- A symbolic/discrete system:

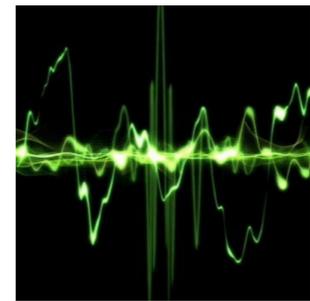
- *table*: 

- *piano*: 

- that pre-dates logic...

Language is special

- Encoded via continuous signals:
 - Sounds
 - Gestures
 - Image (writing)
- Brain encodings are continuous



Why is it hard?

- ambiguity
- variability
- sparsity
- grounding

Ambiguity

*“Finally, a computer that understands you like your mother”
(Ad , 1985)*

1. The computer understands you as well as your mother understands you.
2. The computer understands that you like your mother.
3. The computer understands you as well as it understands your mother.

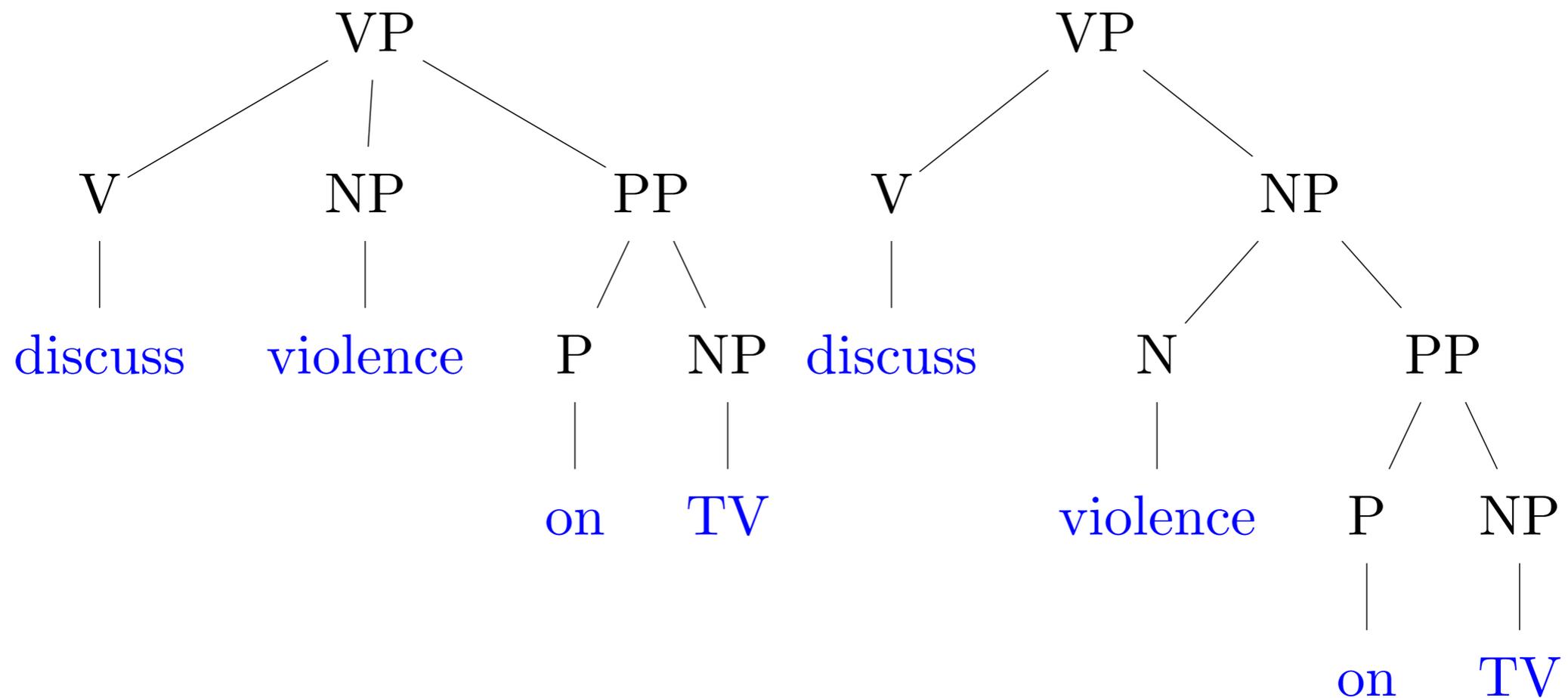
Ambiguity

*“Finally, a computer that understands you like your mother”
(Ad , 1985)*

1. The computer understands you as well as your mother understands you.
2. The computer understands that you like your mother.
3. The computer understands you as well as it understands your mother.

“Finally, a computer that understands your lie cured mother”

Syntactic ambiguity



Even short sentences have hundreds of analyses

Semantic ambiguity

- *There's a river that crosses every town*
- Quantification ambiguity (not syntactic)
- חולצה מטיילת בואדי
- Lexical, syntactic and semantic ambiguity

Ambiguity

- 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*

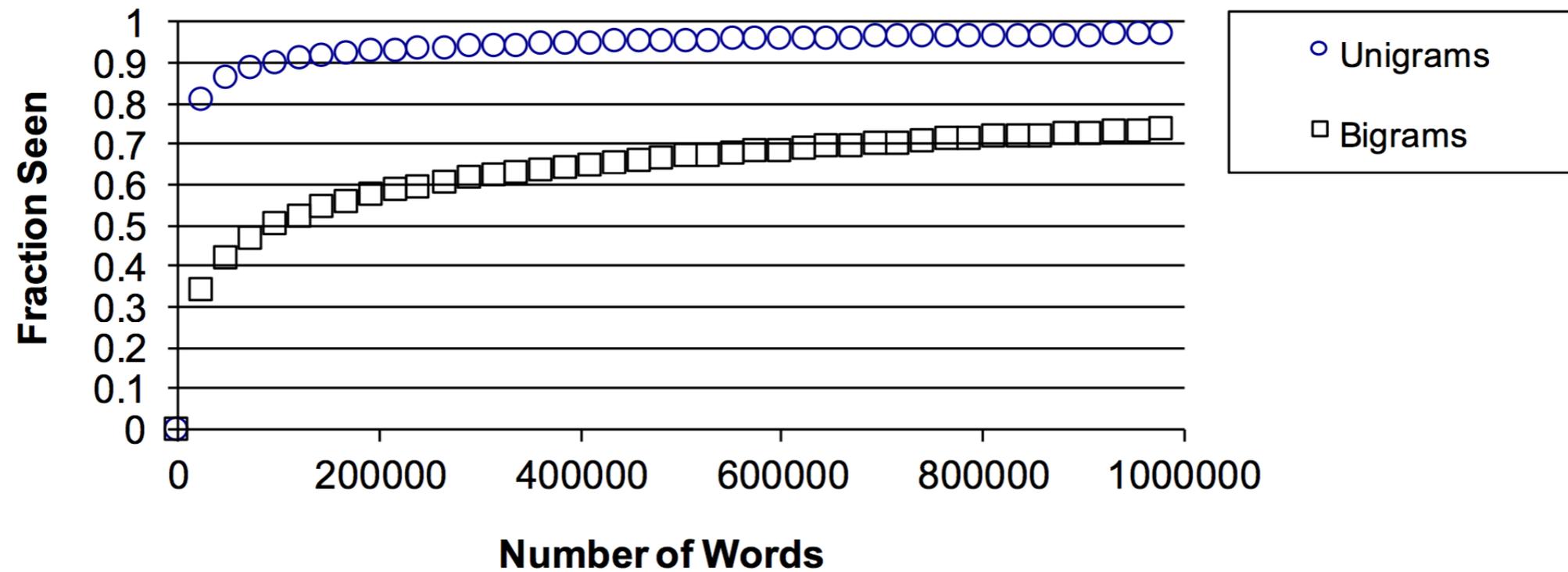
[more](#)

Variability

- Crucial in semantics
 - *Dow ends up 255 points*
 - *Dow climbs 255*
 - *All major stock markets surged*
 - *Dow gains 255 points*
 - *Stock market hits a high record*
 - *The Dow Jones Industrial Average close up 255*



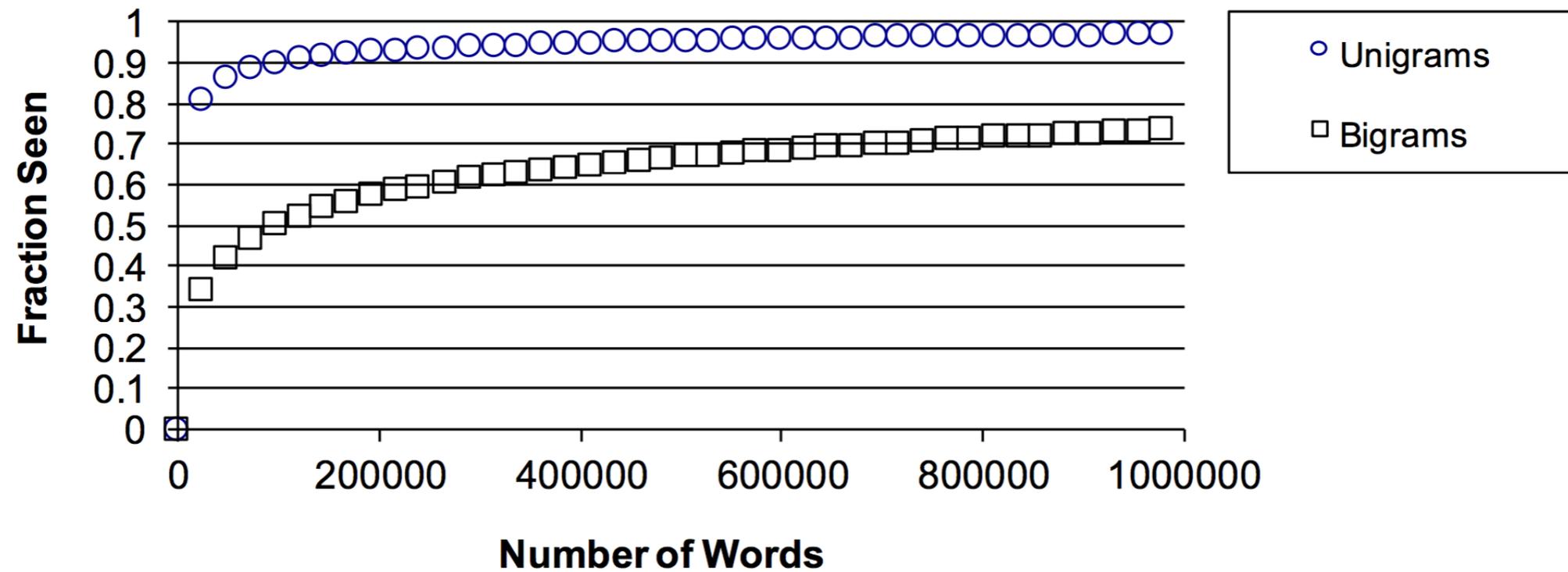
Sparsity



“I ate an apple”

- Unigrams: “I”, “ate”, “an”, “apple”
- Bigrams: “I ate”, “ate an”, “an apple”

Sparsity



“I ate an apple”

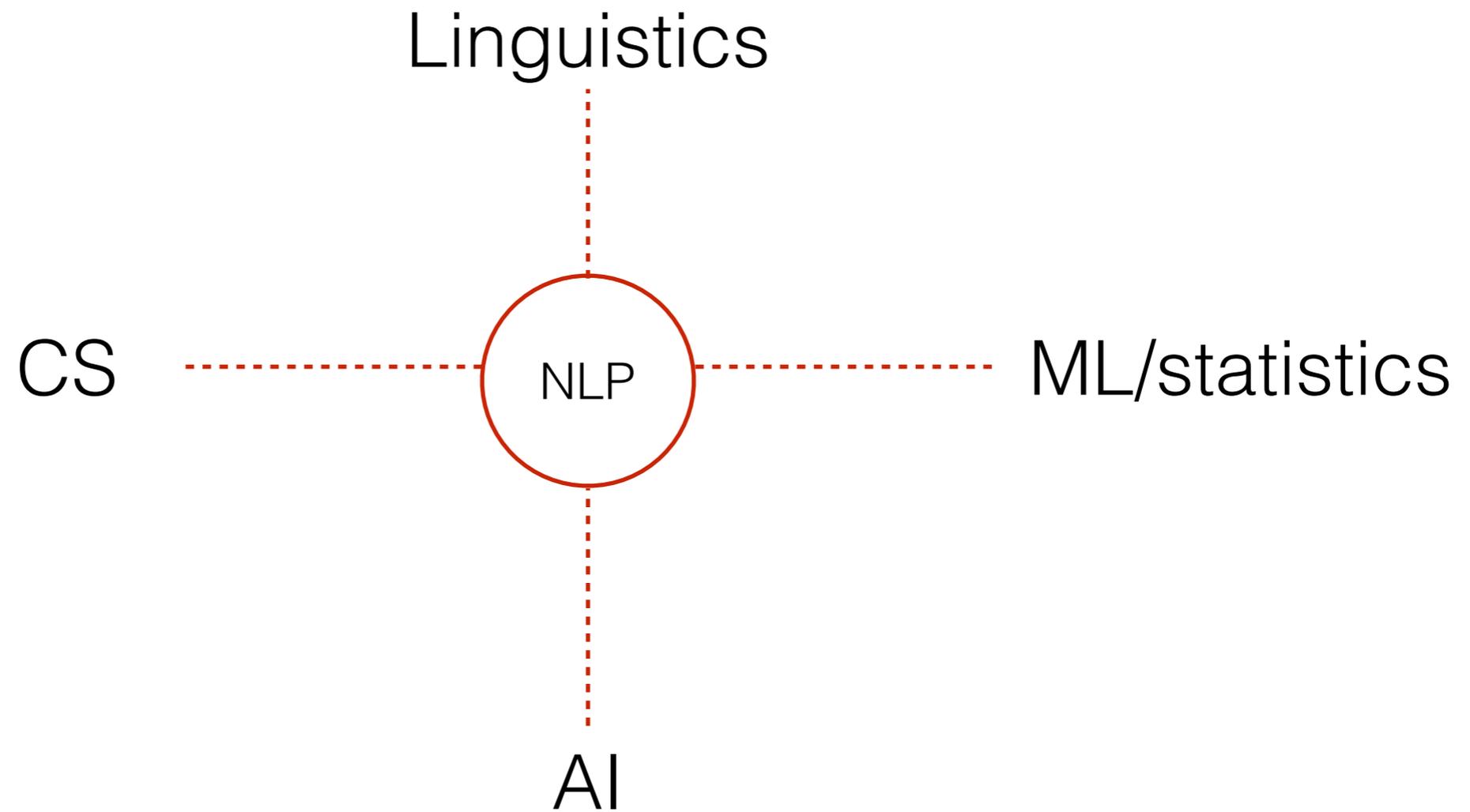
- Unigrams: “I”, “ate”, “an”, “apple”
- Bigrams: “I ate”, “ate an”, “an apple”

Compositionality is key!

Grounding

- Humans do not learn language by observing an endless stream of text.

Interdisciplinary field



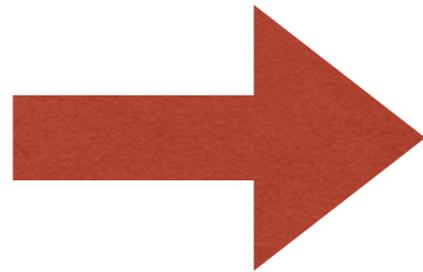
Related fields

- Computational Linguistics
 - Use computational models to learn about language (e.g., meaning change)
- Cognitive Science
 - How does the human brain process language?

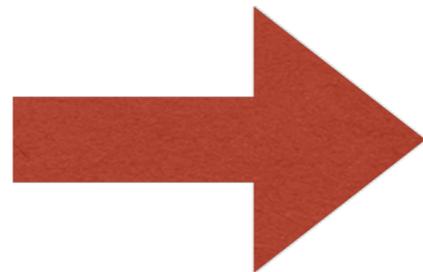
NLP applications

- Text categorization
- Information Extraction
- Search
- Question Answering
- Virtual assistants
- Machine translation
- Summarization
- Reading comprehension

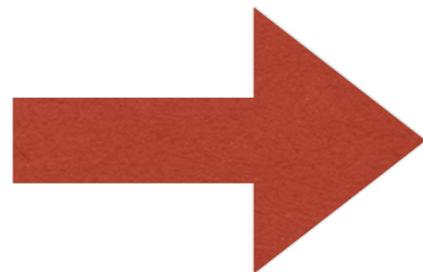
Text categorization



Sports



Politics



Science

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**.

Person	Company	Post	State
Russell T. Lewis	New York Times newspaper	president and general manager	start
Russell T. Lewis	New York Times newspaper	executive vice president	end
Lance R. Primis	New York Times Co.	president and CEO	start

Question answering

What's the capital of Wyoming?



Web

Maps

Shopping

Images

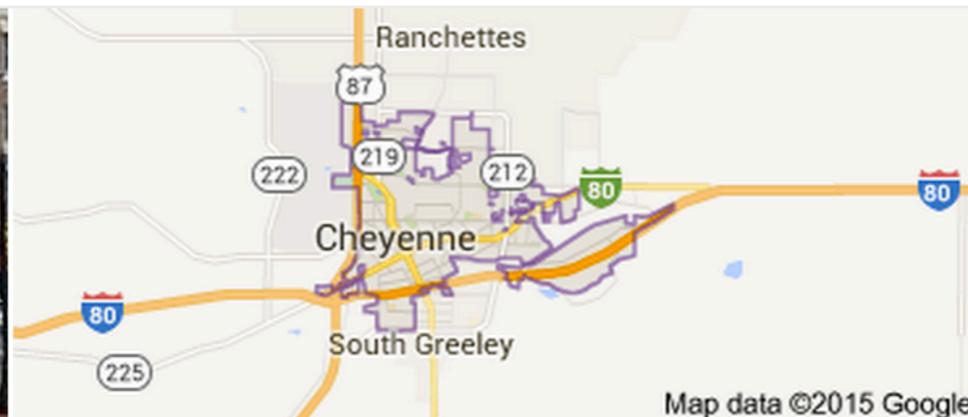
News

More ▾

Search tools

About 984,000 results (0.54 seconds)

Wyoming / Capital



Cheyenne

Question answering

How many US states' capitals are also their largest cities?



Web

Images

News

Shopping

Videos

More ▾

Search tools

About 982,000,000 results (0.67 seconds)

State Capitals and Largest Cities - Infoplease

www.infoplease.com › [United States](#) › [States](#) ▾

State Capitals and Largest Cities. The following table lists the **capital** and **largest city** of every **state** in the **United States**. **State, Capital, Largest city.**

State Capitals and Largest Cities - Fact Monster

www.factmonster.com › [United States](#) › [States](#) ▾ [Fact Monster](#) ▾

State Capitals and Largest Cities. The following table lists the **capital** and **largest city** of every **state** in the **United States**. **State, Capital, Largest city.**

List of capitals in the United States - Wikipedia, the free ...

https://en.wikipedia.org/.../List_of_capitals_in_the_United_Sta... ▾ [Wikipedia](#) ▾

Austin is the largest **state capital** that is not also the **state's largest city**. The Confederate **States of America** had two **capitals** during its existence. The first ... In **many** cases, former **capital** cities of **states** are outside the current **state** borders.
[State capitals](#) - [Insular area capitals](#) - [Former national capitals](#)

Question answering



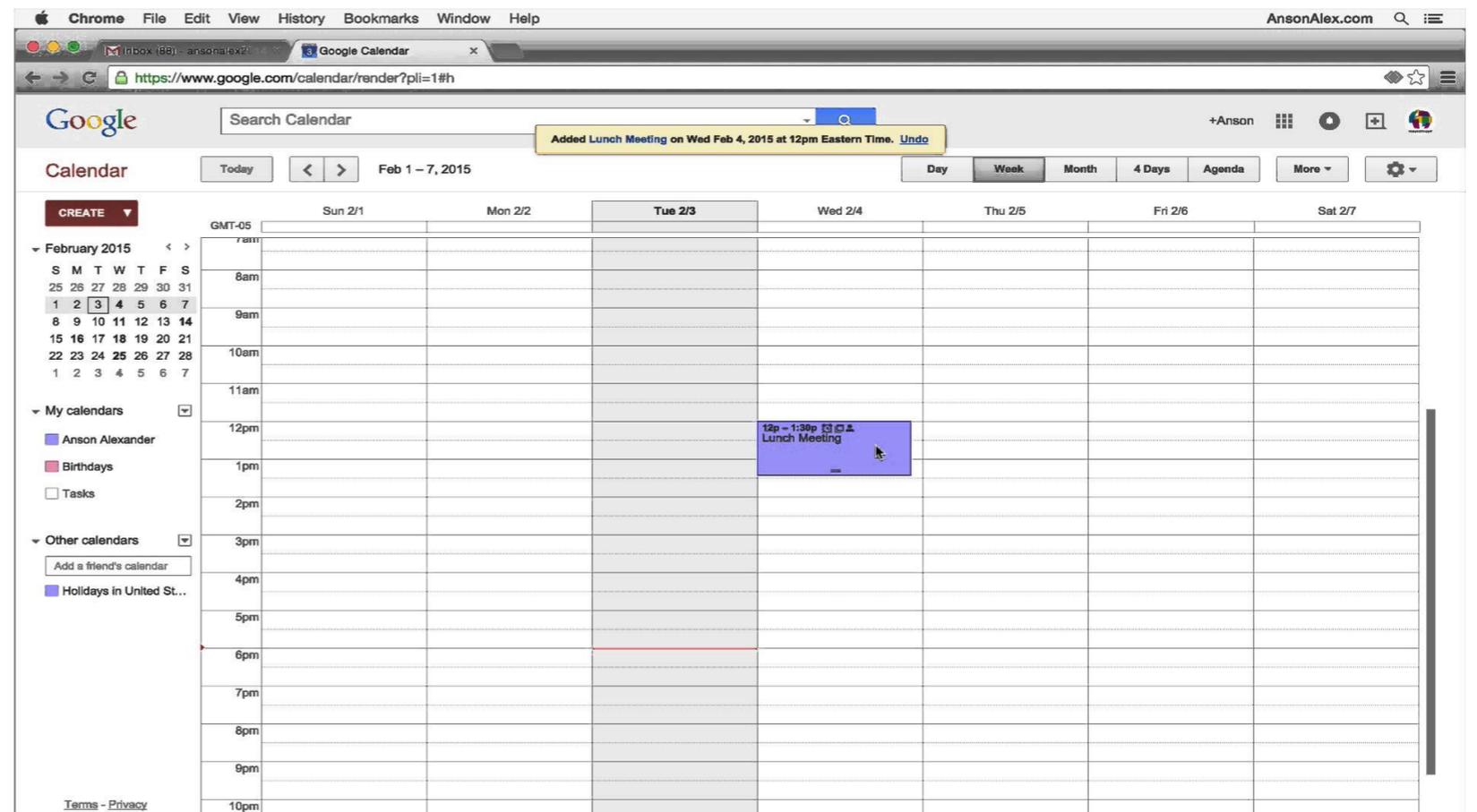
[PR video](#)

Critique

- Douglas Hofstadter: “just a text search algorithm connected to a database, just like Google search. It doesn’t understand what it’s reading.”

Virtual assistants

Move all my Wed meetings in April with John to 5pm



Machine translation

× דונלד טראמפ אולי ניצח בבחירות והושבע לנשיא ארצות הברית, אבל מי שבאמת קובע את המהלכים בבית הלבן הוא האגו השברירי שלו. האגו מכתוב לטראמפ לחתום על צווים נשיאותיים בקצב אש כדי שאיש לא יאמר שאיננו מקיים את הבטחות הבחירות שלו, אבל האגו גם מסבך אותו מידי יום בשקרים חדשים ובוטים שהופכים עתה, בהוראת האגו כמובן, גם למעשים. הצירוף של שני הדברים - הצווים הדרמטיים והפיברוקים העזים - מותיר את ארה"ב הלומה.

394/5000

Donald Trump may have won the election and was sworn in as president of the United States, but who really determines the steps the White House is his fragile ego. Trump ego dictates to sign the decrees rate of fire so that no one would say that we fulfill the promises of his campaign, but also ego complicates it every day new and blatant lies that make now, the teaching of the ego, of course, also acts. The combination of two things - the dramatic and Hfibrokim orders goats - leaving the US in shock.



Suggest an edit

Machine Translation

× דונאלד טראמפ אולי ניצח בבחירות והושבע לנשיא ארצות הברית אבל מי שבאמת קובע את המהלכים בבית הלבן הוא האגו השברירי שלו. האגו מכתוב לטראמפ לחתום על צווים נשיאותיים בקצב אש כדי שאיש לא יאמר שאיננו מקיים את הבטחות הבחירות שלו, אבל האגו גם מסבך אותו מידי יום בשקרים חדשים ובוטים שהופכים עתה, בהוראת האגו כמובן, גם למעשים. הצירוף של שני הדברים - הצווים הדרמטיים והפיברוקים העזים - מותיר את ארצות הברית הלומה.

400/5000

Donald Trump may have won the election and been sworn in as President of the United States, but the one who really sets the pace in the White House is his fragile ego. The ego dictates to Trump to sign presidential orders at a rate of fire so that no one will say that he does not keep his election promises, but the ego also complicates him every day with new, blunt lies that are now, on the ego's orders, The combination of these two things - the dramatic orders and the mighty fibrokes - leaves the United States shocked.



 Suggest an edit

Machine Translation

דונלד טראמפ אולי ניצח בבחירות והושבע לנשיא ארצות הברית אבל מי שבאמת קובע את המהלכים בבית הלבן הוא האגו השברירי שלו. האגו מנתיב לטראמפ לחתום על צווים נשיאותיים בקצב אש כדי שאיש לא יאמר שאיננו מקיים את הבטחות הבחירות שלו, אבל האגו גם מסבך אותו מידי יום בשקרים חדשים ובוטים שהופכים עתה, בהוראת האגו כמובן, גם למעשים. הצירוף של שני הדברים - הצווים הדרמטיים והפיברוקים העזים - מותיר את ארצות הברית הלומה.

×

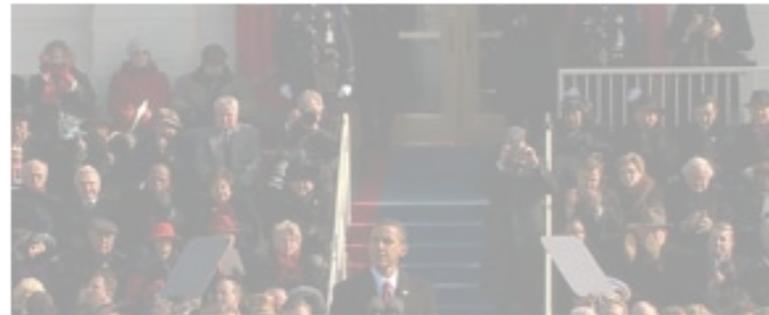
Donald Trump may have won the election and was sworn in as President of the United States, but the one who really sets the pace in the White House is his fragile ego. The ego dictates to Trump to sign presidential decrees at a rate of fire so that no one will say that he does not keep his election promises, but the ego also complicates him every day with new, blatant lies that are now, on the ego's orders, of course, also actions. The combination of these two things - the dramatic orders and the powerful fibrokes - leaves the United States shocked.

☆



Summarization

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.



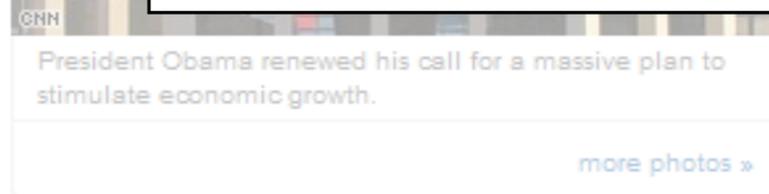
Some inaugural addresses are known for their soaring, inspirational language. Like John F. Kennedy's in 1961: "Ask not what your country can do for you. Ask what you can do for your country."

Obama's address was less stirring, perhaps because it was also more candid and down-to-earth.

"Starting today," the new president said, "we are moving forward."

STORY HIGHLIGHTS

- Obama's address less stirring than others but more candid, analyst says
- Schneider: At a time of crisis, president must be reassuring
- Country has chosen "hope over fear, unity of purpose over ... discord," Obama said
- Obama's speech was a cool speech, not a hot one, Schneider says



President Obama renewed his call for a massive plan to stimulate economic growth.

[more photos »](#)

his first inaugural in 1933, "The only thing wrong with America is that we are afraid. The only thing we have to fear is fear itself." Or Bill Clinton, who took office during the economic crisis of the early 1990s. "There is nothing wrong with America that cannot be fixed by what is right with America," Clinton declared at his first inaugural.

[Obama](#), too, offered reassurance.

"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 arms. 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.

Reading Comprehension

"The rock was still wet. The animal was glistening, like it was still swimming," recalls Hou Xianguang. Hou discovered the unusual fossil while surveying rocks as a paleontology graduate student in 1984, near the Chinese town of Chengjiang. "My teachers always talked about the Burgess Shale animals. It looked like one of them. My hands began to shake." Hou had indeed found a *Naraoia* like those from Canada. However, Hou's animal was 15 million years older than its Canadian relatives.

It can be inferred that Hou Xianguang's "hands began to shake" because he was

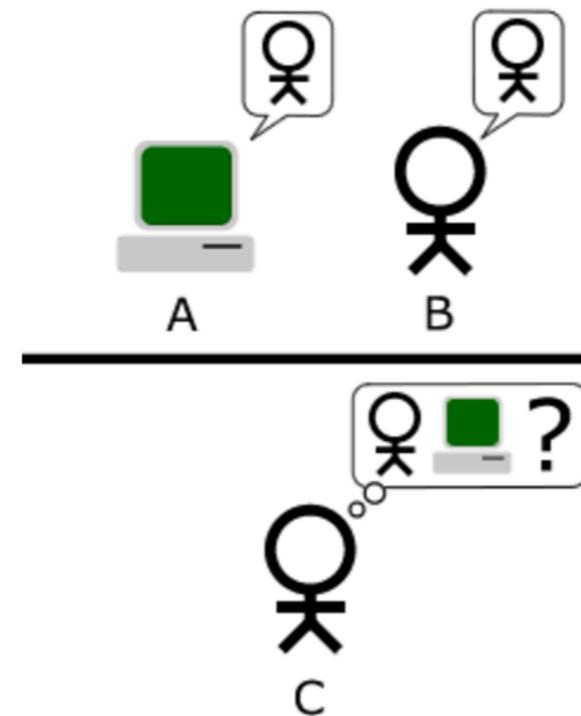
- (A) afraid that he might lose the fossil
- (B) worried about the implications of his finding
- (C) concerned that he might not get credit for his work
- (D) uncertain about the authenticity of the fossil
- (E) excited about the magnitude of his discovery

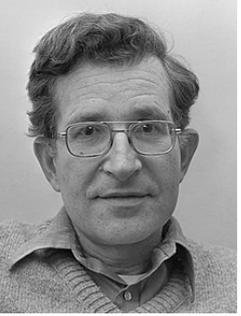


Some history

I propose to consider the question, "Can machines think?" This should begin with definitions of the meaning of the terms "machine" and "think"... If the meaning of the words "machine" and "think" are to be found by examining how they are commonly used... is to be sought in a statistical survey such as a Gallup poll. But this is absurd... I shall replace the question by another, which is closely related to it and is expressed in relatively unambiguous words.

Turing test:



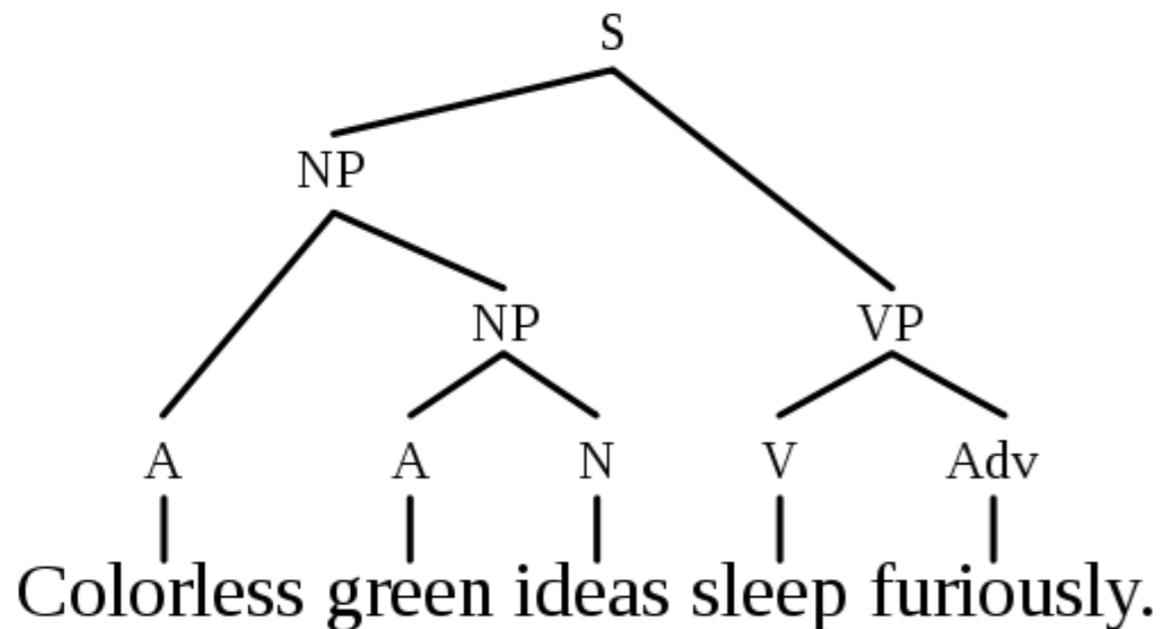


Some history

A. *Colorless green ideas sleep furiously*

B. *Furiously sleep ideas green colorless*

“It is fair to assume that neither sentence (1) nor (2) ... had ever occurred in an English discourse. Hence, in any statistical model for grammaticality, 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, Syntactic structures, 1957)

History: rule-based

- 70s and 80s:
 - Grammars (rules) of English syntax
 - Small domains
 - Substantial engineering effort

$S \rightarrow NP VP$

$NP \rightarrow DET N$

$VP \rightarrow V_T N$

$VP \rightarrow V_I$

$N \rightarrow \text{man}$

$N \rightarrow \text{woman}$

$DET \rightarrow \text{the}$

$DET \rightarrow \text{a}$

$V_T \rightarrow \text{hugged}$

$V_I \rightarrow \text{slept}$



Eliza

Please tell me what's bothering you.

Oh, hello, ELIZA. It's been a long time.

Hi. What seems to be your problem ?

Lately, I've been feeling a little depressed.

And how do you feel right now ?

I feel sad, but relieved to be writing.

Of what does feeling sad remind you ?

Well, I'm an alcoholic, and I've been sober for a long time, but I drank yesterday. And then I decided to stop drinking again. But still that makes me feel sad.

Is it because you are an alcoholic that you came to me ?

Empirical revolution

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

Jelinek (?), 1988

When I look at an article in Russian, I say: 'This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode.'

Weaver, 1955

Of course, we must not go overboard and mistakenly conclude that the successes of statistical NLP render linguistics irrelevant (rash statements to this effect have been made in the past, e.g., the notorious remark, “Every time I fire a linguist, my performance goes up”). The information and insight that linguists, psychologists, and others have gathered about language is invaluable in creating high-performance broad domain language understanding systems; for instance, in the speech recognition setting described above, better understanding of language structure can lead to better language models.” (Lilian Lee, 2001)

Empirical Revolution

- 1990: corpus-based statistical methods
- 2000s: rich structural formalisms
- 2010s: representation learning with neural nets

Class goals

- Appreciate the challenges of NLP
- Learn about NLP problems and their solution
- Learn general purpose methods
- Be able to read modern research in NLP

Disclaimer

- This class borrows from NLP classes around the world
 - Statistical NLP by Michael Collins
 - Stanford 224n
 - Others
- You are welcome to watch videos, read those if you feel more comfortable with that
- However - **slides are not self-contained**. You need to be in class to fully understand their meaning
- Classes might be video-taped?

Plan

- Technical foci: structured prediction and deep learning
- **Class 1:** introduction + word embeddings
- **Class 2:** Word embeddings
- **Class 3:** Language models, contextualized word representations
- **Class 4:** RNNs
- **Class 5:** Tagging, log-linear models
- **Class 6:** Globally-normalized models, deep learning for tagging

Plan

- **Class 7:** Introduction to parsing, CKY
- **Class 8:** Lexical PCFG
- **Class 9:** Deep syntactic parsing
- **Class 10:** Semantic parsing, sequence-to-sequence models
- **Class 11:** Projects
- **Class 12:** Sequence-to-sequence models

Prerequisites

- Introduction to machine learning
- Proficiency in programming (some assignments require python)
- Implied from intro to ML:
 - Probability
 - Linear algebra
 - Calculus
 - Algorithms

Assignments (tentative)

- Every two classes. Tentative plan:
 - Assignment 1: Word embeddings
 - Assignment 2: Language modeling
 - Assignment 3: Sequence tagging
 - Assignment 4: Deep learning
 - Assignment 5: Syntax

Assignments

- Five homework assignments
- All involve coding
- Submit in pairs or triplets
- Submit by e-mail to the 2 TAs:
 - E-mail title: Assignment <number>: <firstname1>
<lastname1> <id1> <firstname2> <lastname2> <id2>
- Some simple written questions

Administration

- **Instructor:** Jonathan Berant
- **TAs:** Ohad Rubin, Omri Koshorek
- **Time:** Tue, 13-16
- **Place:** Dan David 001
- **Office hours:** coordinate by e-mail
- **Website:** https://www.cs.tau.ac.il/~joberant/teaching/advanced_nlp_spring_2019/index.html
- **Forum:** <http://moodle.tau.ac.il/course/view.php?id=368307701>
- **Grades:** 50% final project, 50% homework assignments

Final project

- Groups of 2-3 people (best to just have same team as for home assignments)
- There will be a default project - TBD
- Or you can do a research projects
 - Choose a research paper, implement it, extend it
 - Build a model that does something you care about
 - Define a research problem and work on it
- If you are taking Prof. Globerson's class (advance ML) it is advised that you have one project for both classes. Standard for such project will be higher. **All members of team must be enrolled in both classes.**

Final project

- Research projects:
 - Decide on project by roughly week 10.
 - Get 1-bit approval
 - Present in last class
- 2-3 top-tier papers result from the class per year.
- See class website for examples

Textbooks

- Martin and Jurafsky:
 - Speech and Language Processing: An Introduction to Natural Language Processing
- Schutze and Manning:
 - Statistical Foundations of NLP
- Noah Smith:
 - Linguistic structure prediction
- Yoav Goldberg:
 - Neural Network Methods for Natural Language Processing