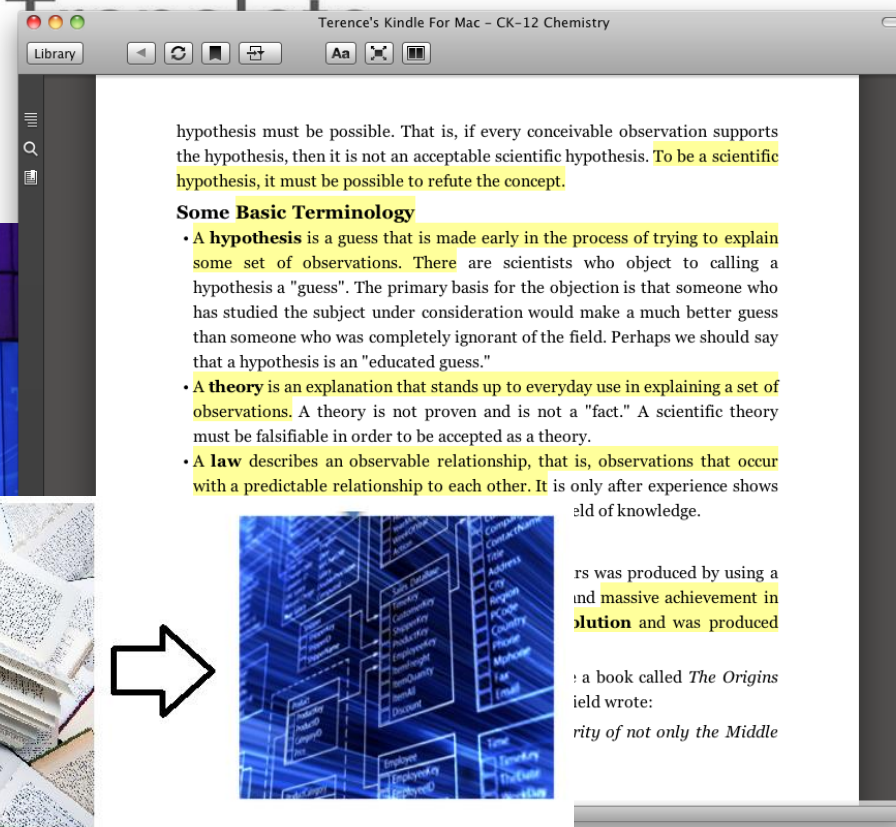
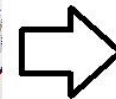
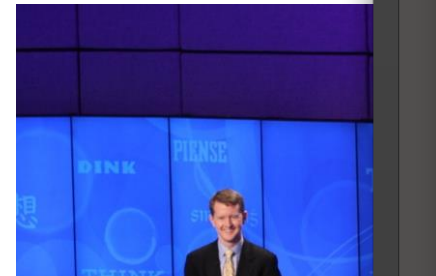


# Natural Language Processing

## A Brief Introduction

# Natural Language Processing is now part of everyday life

- Information Retrieval
- Machine Translation
- Question Answering
- Summarization



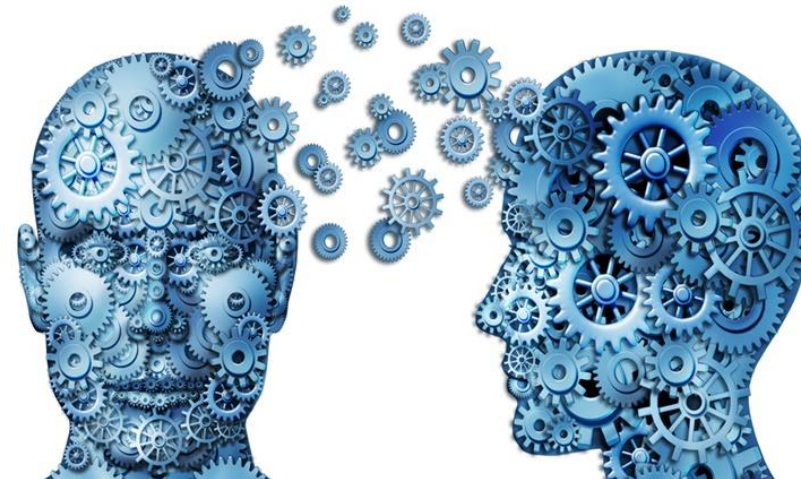
# Natural Language Processing: Dual Nature

---

1. Tools for processing language



2. Models of human linguistic  
behavior and acquisition



# Brief History of NLP

---

- **70s-80s:** rule-based methods:
  - Deeper models of syntax and semantics
  - Toy domains, models generally did not scale well
- **90s:** corpus-based methods, empirical evaluation
  - Deep linguistic analysis often traded for robust methods
- **2000s:** sophisticated machine learning methods, scalability, real-world
- **Present:** much interest in semantics, text-to-text generation and deep learning



# We've only just Scratched the Surface

- Despite being 50 years old, most important questions in the field are still open
- Most classic questions still unanswered: (some were replaced by other questions...)
  - Machine Translation
  - Universal semantic representation (content rather than form)
  - Question answering: answer natural language queries
- In the course we will cover both classic material and very recent work



# NLP: Intellectually Stimulating Debates

---

- *It must be recognized that the notion “probability of a sentence” is an entirely useless one, under any known interpretation of this term.*  
Noam Chomsky, 1969
- *Whenever I fire a linguist our system performance improves.*  
Frederick Jelinek, 1988 [a founder of modern automatic speech recognition and a prominent computational linguist]



# Machine Translation

- Translate text from one language to another
- Statistical Machine Translation mostly uses parallel corpora as training data
- The Canadian and EU Parliament proceedings have been very useful in this respect

Arabic	English	Spanish
ووصولاً إلى هذه الغاية، عدد أربع نقاط أساسية، وهي بالتحديد، دعم المجلس القوي للسعي إلى إيجاد حل إقليمي، والمشاركة النشطة لـ دول الإقليم في تحقيق هذا الهدف، والإشارة الواضحة إلى أن الحل الحقيقي ينبغي أن يفي بالحد الأدنى من متطلبات جميع الأطراف، لا أن يحقق المطالب القصوى لأي منهم، والتوصل إلى اتفاق يكون عنصراً أساسياً في إطار نسوية أوسع تشمل المنطقة ككل.	To that end, he listed four benchmarks, <b>namely</b> , the firm support of the <b>Council</b> for the search for a regional solution, the active participation of the <b>States</b> of the region in achieving that goal, the clear indication that a true solution should meet the minimum requirements of all parties, but the maximum demands of none, and an agreement that was firmly embedded in the context of a broader settlement encompassing the region as a whole.	A tal fin , enunció cuatro condiciones , <b>a saber</b> : el firme apoyo del <b>Consejo</b> a los esfuerzos por lograr una solución regional , la participación activa de los <b>Estados</b> de la región en la realización de ese objetivo , una clara indicación de que toda solución verdadera debía satisfacer los requisitos mínimos de todas las partes , pero no las exigencias máximas de cualquiera de ellas , y un acuerdo firmemente arraigado en el contexto de una solución más amplia que abarcase a la región en su totalidad .

# Machine Translation

---

- The model learns by learning correspondences between fragments of the text on both sides
- Challenges:
  - What fragments to take?
  - How to combine them?
  - How to make efficient?
- State of the art: far from perfect but reasonable as assistive technology

**Le Monde.fr**

La Bourse de Shanghai dégringolait de plus de 6 % mardi 25 août à l'ouverture, après s'être déjà effondrée de presque 8,5 % la veille, dans un marché affolé par l'affaiblissement persistant de l'économie chinoise et miné par des inquiétudes sur la conjoncture mondiale.

Dans les premiers échanges, l'indice composite chutait de 6,41 % soit 205,78 points à 3 004.13 points. La Bourse de Shenzhen plongeait quant à elle de

The Shanghai Stock Exchange tumbled more than 6% Tuesday, August 25 at the opening, having already collapsed by almost 8.5% yesterday, in a panicked market the persistent weakening of the Chinese economy and undermined by concerns about the global economy.

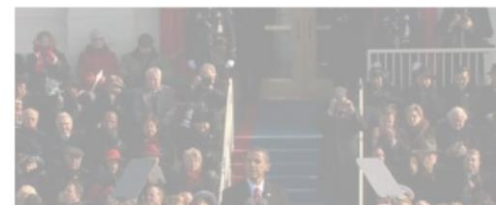
In early trade, the composite index fell by 6.41% or 205.78 points to 3 004.13 points. The Shenzhen Stock Exchange dived for its 6.97% to 1 751.28 points. The Hong Kong Stock Exchange, meanwhile, opened down 0.67%.



# Summarization

- Condense text into a human-readable summary
- Two main types:
  - Extractive
  - Abstractive
- Very goal-dependent
  - Difficult to evaluate

WASHINGTON (CNN) -- President Obama's inaugural address was cooler, more measured and reassured 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 it was also more candid and down-to-earth.

#### 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 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.

# Information Extraction

---

- Goal: map text to database entries

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.

- State of the art:
  - Perhaps 80% accuracy for multi-sentence templates
  - 90%+ for single easy fields
  - But remember: information is redundant, so we can often use this redundancy to denoise

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

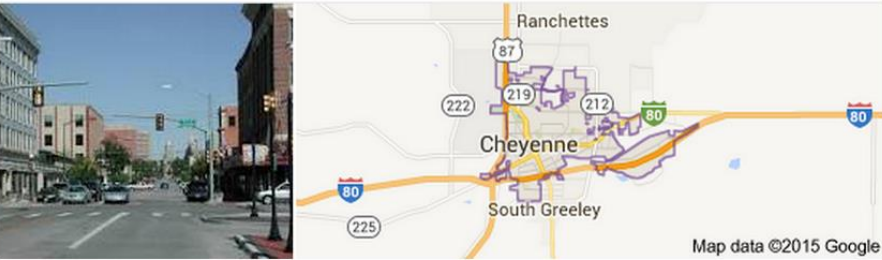
- From a knowledge base or through text mining
- More than search

What's the capital of Wyoming?

Web Maps Shopping Images News More Search tools

About 984,000 results (0.54 seconds)

Wyoming / Capital



Cheyenne

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](http://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](http://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...](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

# Question Answering

Google

best tourist attractions in jerusalem

All Images Maps News Videos More Settings Tools

About 583,000 results (1.17 seconds)





**Best Tours in Jerusalem - By Mazada Tours (Since 1981)**  
[www.mazadatours.com/Best\\_Tours/in\\_Jerusalem](#)  
With our special day tours you can expect to receive the **best** service available. Instant Confirmation · 5,000 Day Tour Activities · Fast&Easy Online Booking... Services: Most Popular Itineraries, Pick up From your Hotel, Satisfaction ...  
Jerusalem Old&New ... Massada &DeadSea ...  
The Western Coast D... Holy-Land Sites Day T...

**7 Different Jerusalem Tours - Culture - Politics - History**  
[www.toursinenglish.com/Jerusalem-Tours](#)  
Learn In-Depth About the Situation in East & West Jerusalem - Must Do ... First Tours to West Bank · Not to be missed · 10 Years of Experience · Alt... Jerusalem Old City Tour · Jerusalem All Day · Jerusalem East & West · W...

**Jerusalem Tour - See The Top Sights - tripadvisor.com**  
[www.tripadvisor.com/sightseeing](#)  
Read reviews and view photos. Book a **Jerusalem sightseeing tour!**  
Types: Tours, Attraction Tickets, Day Trips, Airport Transfers, Excursions

**10 Best Things To Do In Jerusalem - A Complete Guide to Jerusalem**  
[www.visitcity.com/jerusalem/travel\\_guide](#)  
Create Your Personal Travel Guide to **Jerusalem**. Itineraries, Maps & Att...

Jerusalem / Top sights

			
Church of the Holy Sepulchre Circa-4th-century Christian chapel	Western Wall Ancient sacred site for Jewish	Temple Mount Hilltop compound of religious sites	Dome of the Rock Islamic shrine housing

Google

best tourist attractions outside jerusalem

All Maps Images Videos News More Settings Tools

Tourist attractions > outside > Jerusalem

						
Tower of David	Mamilla Mall	Mount Zion	Calvary	Walls of Jerusalem	Montefiore Windmill	Billy Rose Sculpture garden

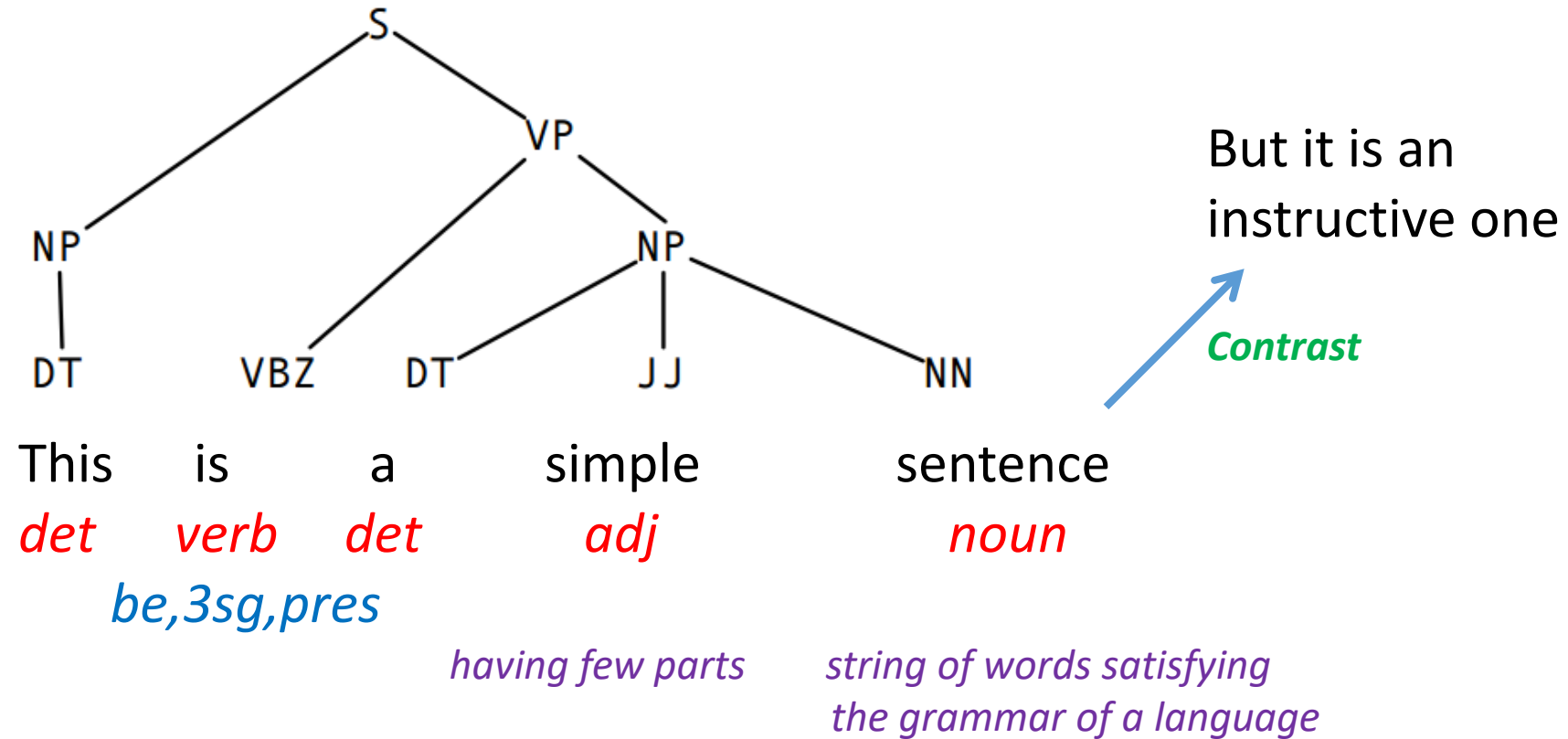
# Layers of Linguistic Representation

---

- Language consists of many levels of structure
- Humans fluently integrate all of these in producing and understanding language
- Some basic representations:

This	is	a	simple	sentence
<i>det</i>	<i>verb</i>	<i>det</i>	<i>adj</i>	<i>noun</i>
<i>be,3sg,pres</i>			<i>having few parts</i>	<i>string of words satisfying the grammar of a language</i>

# Layers of Linguistic Representation





# Key Challenges: Ambiguity

---

- Language has much underlying structure, which is ambiguously expressed
- Ambiguity is at all levels:
  - words can sound the same but not mean the same (bank vs. bank)
  - morphemes can express many different meanings
    - התפעל can be reflexive התרחץ, הסתפר, התפטר but can also have other meanings התפלפל, התכתב etc.
    - The suffix 's' in English can be a plural noun inflection or a singular verb inflection
  - Sentences may look the same, but have different meanings
    - The researcher watched a crocodile with a telescope
- Alongside ambiguity, there is also redundancy: many different ways to say the same thing

# Ambiguity in Headlines

---

- Some real headlines:
  - 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



# Dark Ambiguities

---

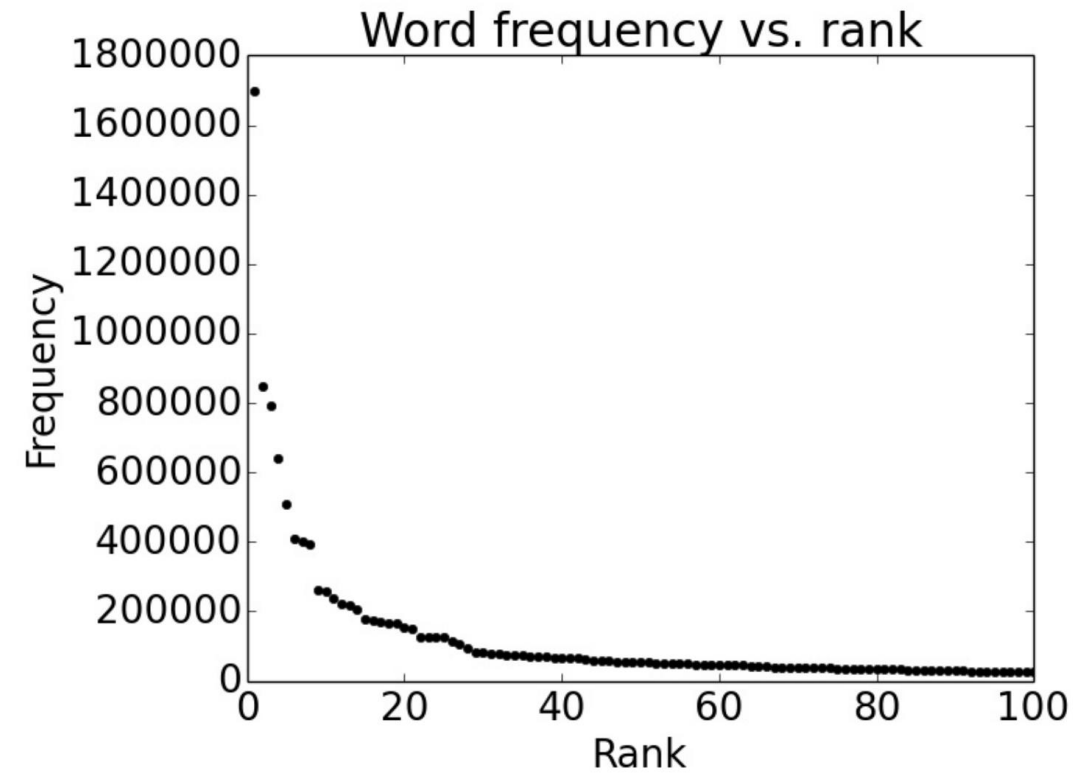
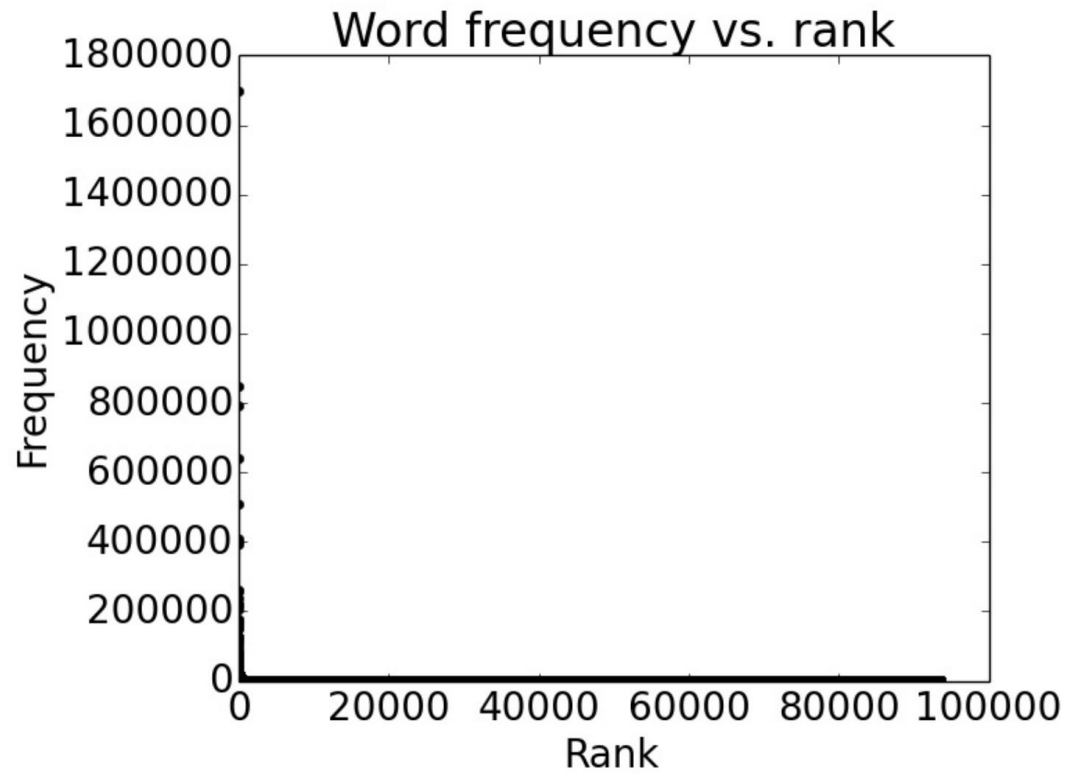
- In fact, ambiguity is much worse than we think: most structurally permitted analyses are so bad that it's hard to think about them

[Rock and Roll] concerts  
**OR**  
Rock and [Roll concerts]

- But can't we overcome this if see enough examples of "Rock and Roll"?

# Sparsity

---



# Sparsity

- Zipf's law:

$$\log(f) \log(r) = k$$

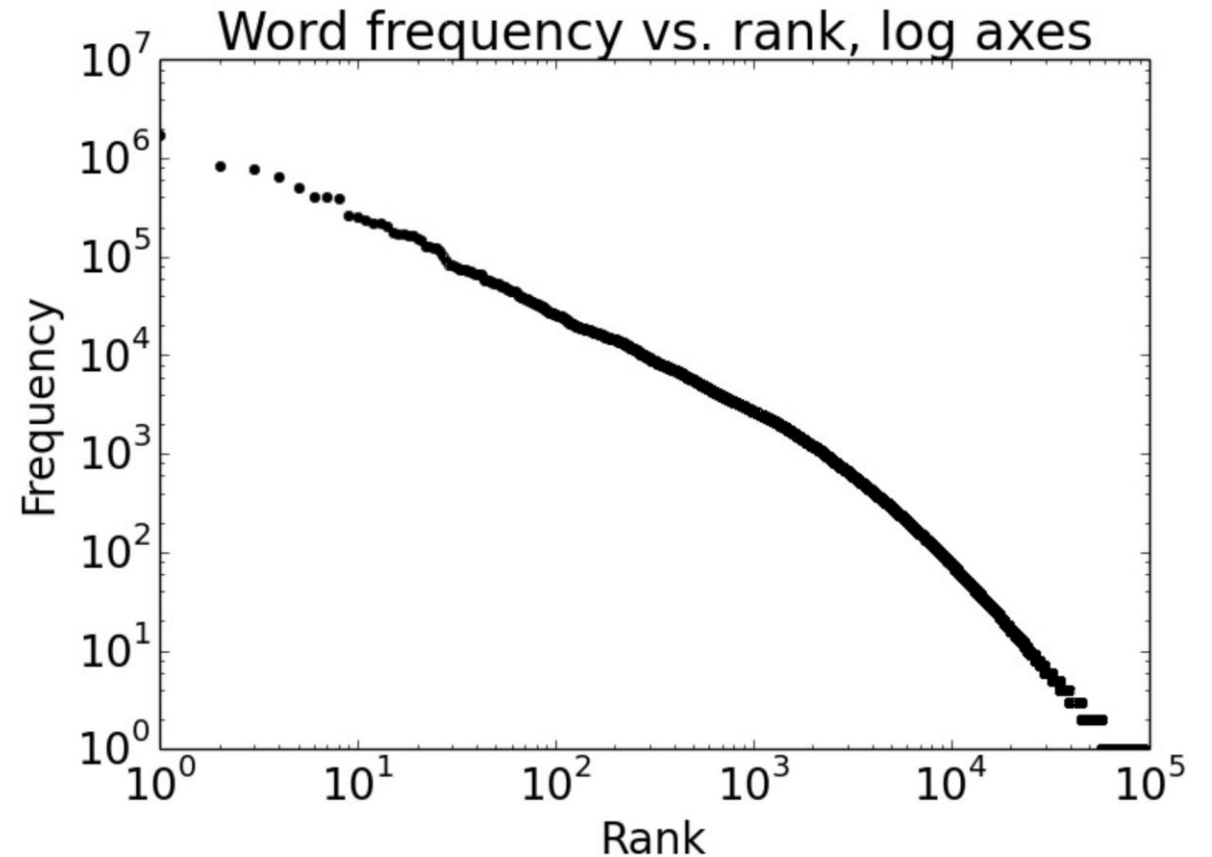
$f$  = frequency of a word

$r$  = rank of a word (if sorted by frequency)

$k$  = a constant

→ *Sparsity is always an issue*

→ *True for many other phenomena too*



# Many Rules, Many Exceptions

- Think of examples in English:
  - Past tense inflection (regular verbs:  $V \rightarrow V+ed$ , but many irregular)
  - “I haven’t a clue” is fine, but “I haven’t a look” is not
  - “The tour begins at 14:40 today” (in the future), but “He is not home at 17” is habitual
    - You should say “He will not be home at 17” for a prediction
  - “River Thames” but “Mississippi River”



# What will we cover?

---

## Natural Language Processing

- Language models
- Document classification
- POS tagging
- Sentiment Analysis
- Syntax and Parsing: grammar-based methods, graph-based methods, transition-based methods
- Sentence Semantics
- Information Extraction
- Annotation and Resources

## Learning

- Probabilistic Modeling: Generative and discriminative models
- Naïve Bayes classifiers
- Log-linear classifiers
- Feed-forward Networks
- Markov Models
- Hidden Markov Models
- Conditional Random Fields
- Structured Prediction (e.g., Structured Perceptron)
- Recurrent Neural Networks

# Useful Resources

---

- Coursera courses:

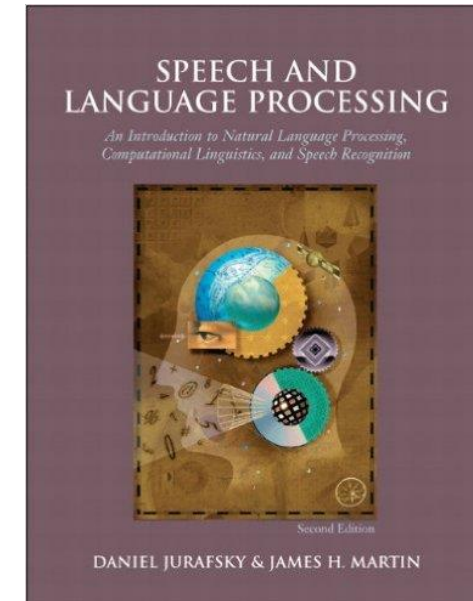
**From Stanford:**

<https://www.coursera.org/course/nlp>

**From Columbia:**

<https://www.coursera.org/course/nlangp>

- Speech and Language Processing,  
2nd Edition, Jurafsky and Martin, 2008
  - Draft version of 3<sup>rd</sup> edition is  
on Dan Jurafsky's website



# Useful Resources

---

- ACL Anthology
  - Leading conferences are ACL, NAACL, EACL, EMNLP, CoNLL
  - Leading journals are TACL and Computational Linguistics
- Google Scholar