Natural Language Processing

Lecture 1: Introduction

Course's Email/Staff

- **All course-related emails should be directed here: Nlp.huji.2022@gmail.com**

- Lecturer: Omri Abend (omri.abend@mail.huji.ac.il)
  - Office hour: Monday at 11am, Computer Science A527

- Tsar: Eitan Wagner (eitan.wagner@mail.huji.ac.il)
  - Office hour: Monday at 12pm, Computer Science A503

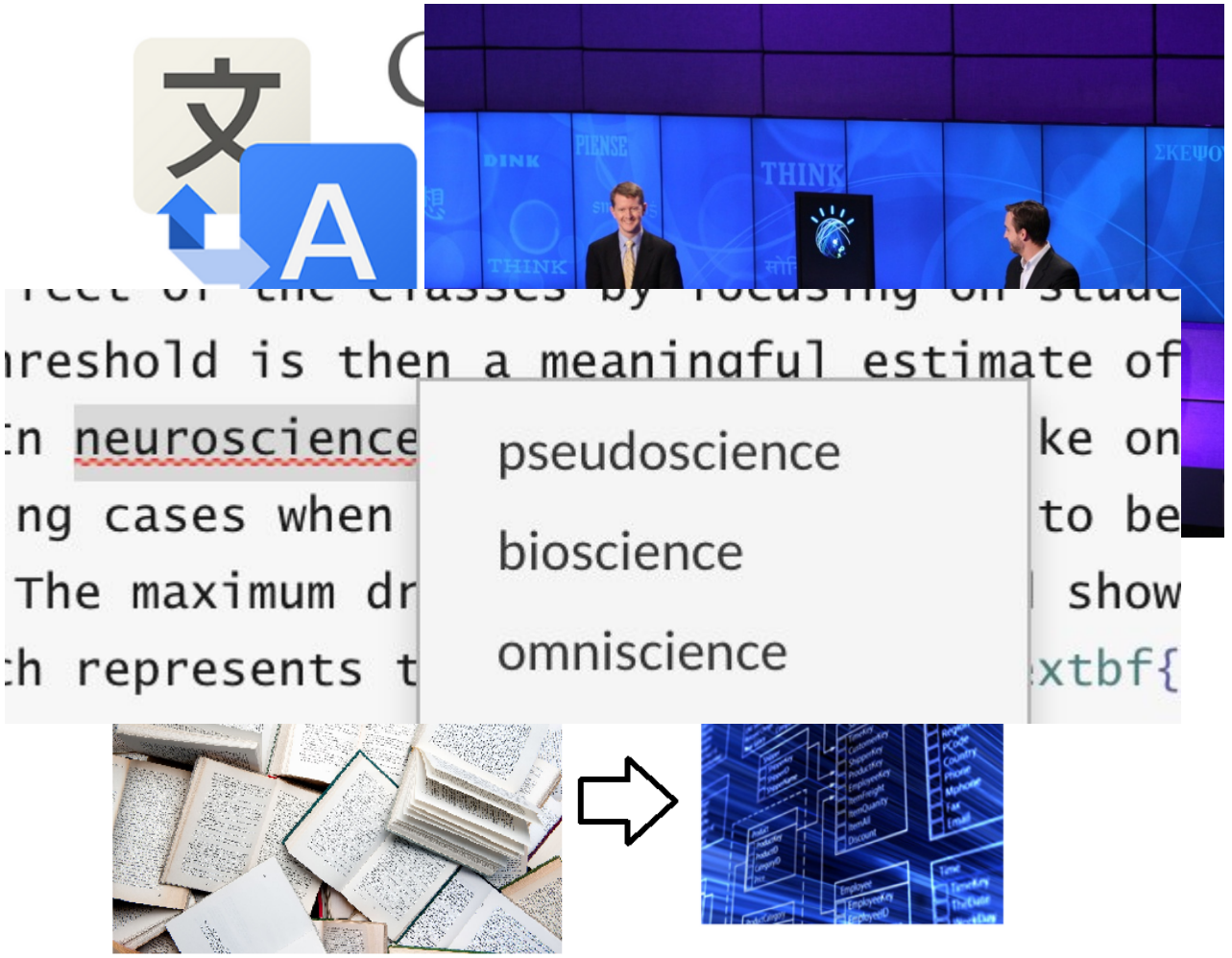
Regulations

The course evaluation will be composed on the following components:

1. Exercises (5 in total) will account for 30% of the final grade
   - The exercises will include both theoretical assignments and programming assignments
   - The assignments will be done in pairs (students who cannot find a pair are asked to contact Eitan)
2. The final exam (standard classroom exam) will account for 70% of the final grade
3. Students will be required to attend in at least 75% of the lecturers. We will use a QR code for each lecture, that you will need to scan.
   - If you have technical difficulties, please send an email to Eitan

Natural Language Processing is now part of everyday life

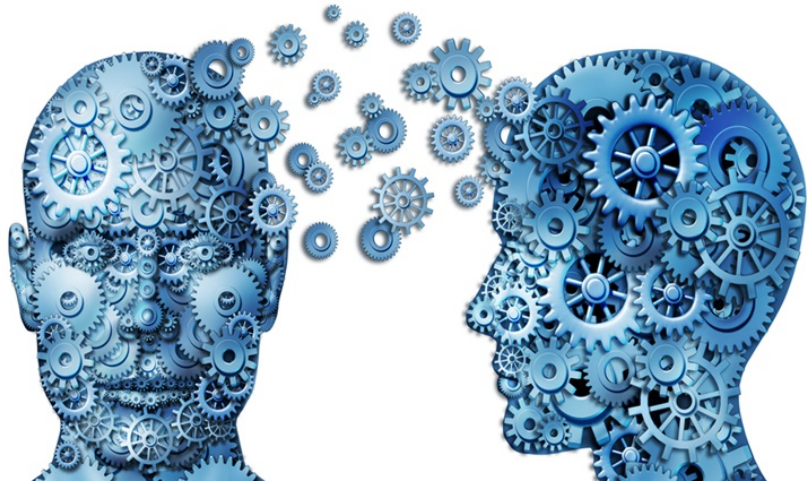
- Information Retrieval
- Machine Translation
- Question Answering
- Summarization
- Grammatical Error Correction



Natural Language Processing: Dual Nature

1. Technology 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

Speedy Progress in Recent Years, but..

- Thanks to new technology (mostly ML), speedy progress in recent years in age-old tasks
  - E.g., think of Google Translate 5 years ago and today
- Still, many important questions are open
- Some classic questions still not fully answered:
  - NLP tools that can generalize to any language
  - Integrating discrete reasoning (e.g., database querying, mathematics) with language understanding
  - Avoid producing blatantly wrong answers (e.g., report failure instead)
- In the course we will cover both classic material and very recent work, but mostly focus on the foundations of the field

Machine Translation

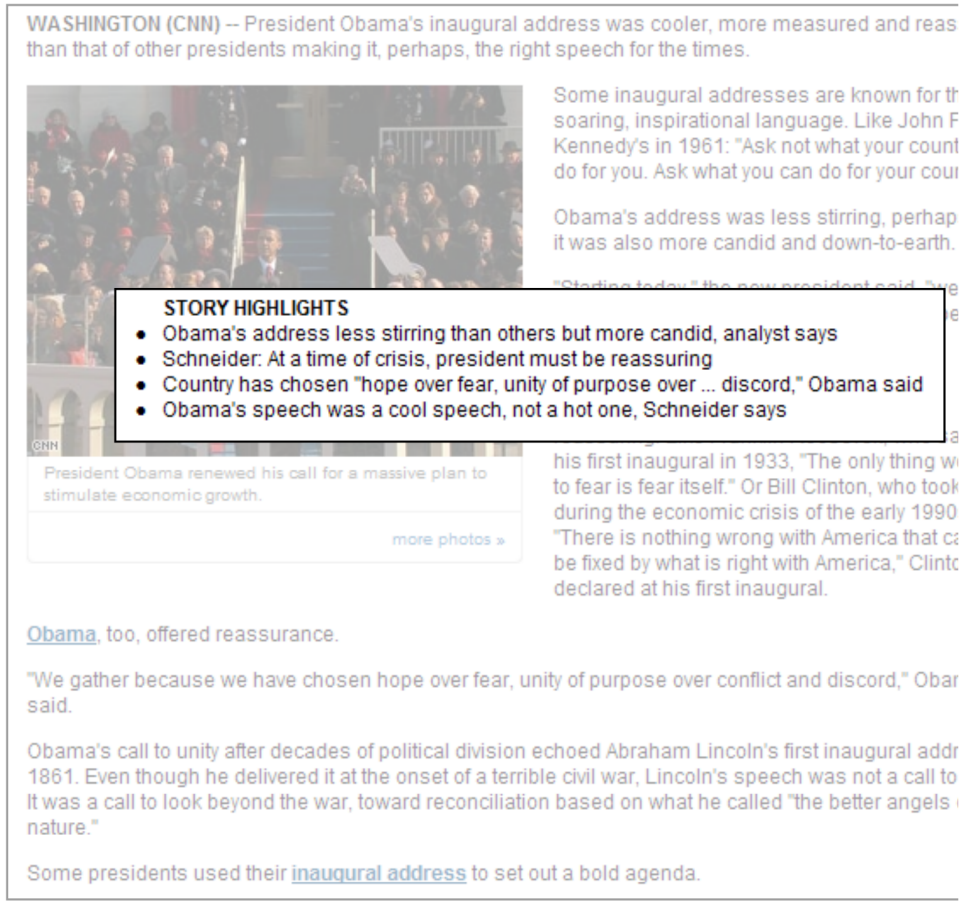
- Translate text from one language to another
- Statistical Machine Translation mostly uses parallel corpora as training data
- The Canadian and EU Parliament proceedings have been very useful in this respect

Machine Translation

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

Summarization

- Condense text into a human-readable summary
- Two main types:
  - Extractive
  - Abstractive
- Very goal-dependent
  - Difficult to define and evaluate
- Getting better, but blatant errors are still pervasive
  - E.g., “hallucinated” entities


WASHINGTON (CNN) -- President Obama's inaugural address was cooler, more measured and reas
than that of other presidents making it, perhaps, the right speech for the times.

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

Obama's address was less stirring, perhap
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

CNN

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

more photos »

his first inaugural in 1933, "The only thing w
to fear is fear itself." Or Bill Clinton, who took
during the economic crisis of the early 1990
"There is nothing wrong with America that ca
be fixed by what is right with America," Clinto
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," Obar
said.

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

Some presidents used their inaugural address to set out a bold agenda.

Summarization: Hallucinations (example)

DOCUMENT: The Richmond Park and North Kingston MP said **he was "honoured" after winning** 70% of the 9,227 votes cast using an online primary system.
He beat London Assembly Member Andrew Boff, MEP Syed Kamall and London's deputy mayor for crime and policing Stephen Greenhalgh.
Mr Goldsmith's main rival is likely to be **Labour's Sadiq Khan**. *(2 sentences with 59 words are abbreviated here.)*
Mr Goldsmith, who was **the favourite for the Tory nomination**, balloted his constituents earlier this year to seek permission to stand.
At the very point of **his entry into the race for London mayor**, **Zac Goldsmith**'s decision revealed two big characteristics. *(5 sentences with 108 words are abbreviated here.)*
Mr Goldsmith - who first entered Parliament in 2010 - told the BBC's Daily Politics that he hoped his environmental record would appeal to Green and Lib Dem voters and he also hoped to "reach out" to **UKIP** supporters frustrated with politics as usual and the UK's relationship with the EU.
Zac Goldsmith Born in 1975, educated at Eton and the Cambridge Centre for Sixth-form Studies *(5 sentences with 76 words are abbreviated here.)*
Mr Goldsmith, who has confirmed he would stand down from Parliament if he became mayor, triggering a by-election, said he wanted to build on **current mayor Boris Johnson**'s achievements. *(3 sentences with 117 words are abbreviated here.)*
Both **Mr Khan and Mr Goldsmith** oppose a new runway at Heathrow airport, a fact described by the British Chambers of Commerce as "depressing". *(1 sentences with 31 words is abbreviated here.)*
Current mayor Boris Johnson will step down next year after two terms in office. He is also currently the MP for Uxbridge and South Ruislip, having been returned to Parliament in May.
Some **Conservatives** have called for an inquiry into **the mayoral election process** after only 9,227 people voted - compared with a 87,884 turnout for the Labour contest. *(4 sentences with 121 words are abbreviated here.)*

PTGEN UKIP leader Nigel Goldsmith has been elected as the new mayor of London to elect a new Conservative MP.

Information Extraction (IE)

- Goal: map text to database entries

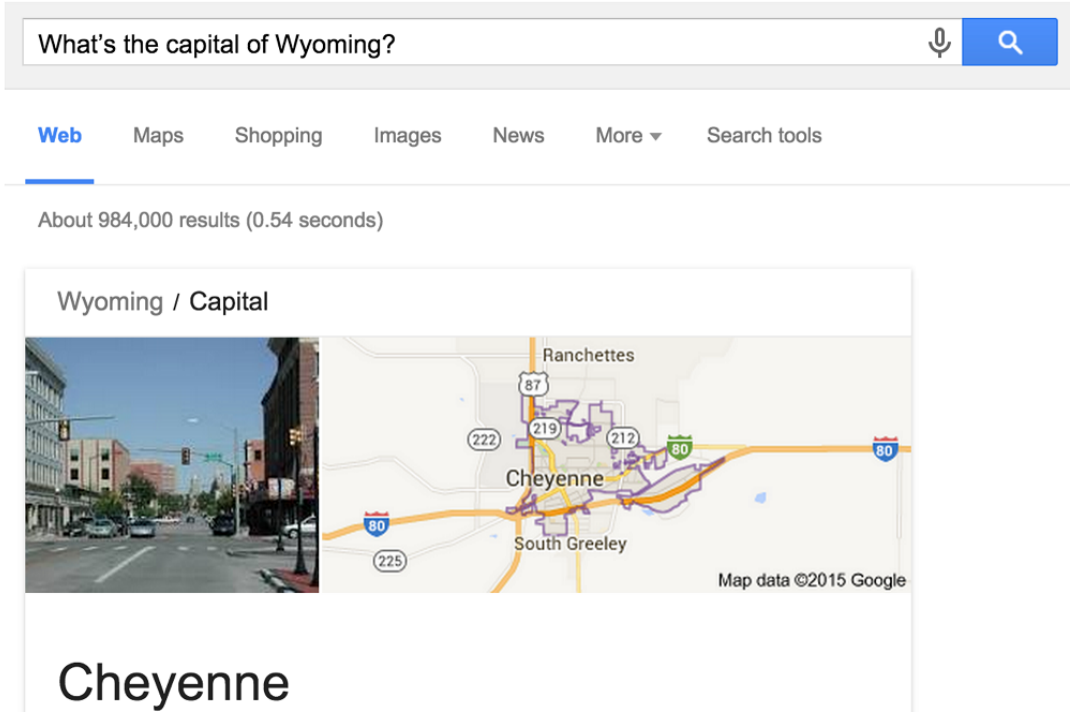
New York Times Co. named Russell T. Lewis, 45, president and general manager of its flagship New York Times newspaper, responsible for all business-side activities. He was executive vice president and deputy general manager. He succeeds Lance R. Primis, who in September was named president and chief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

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

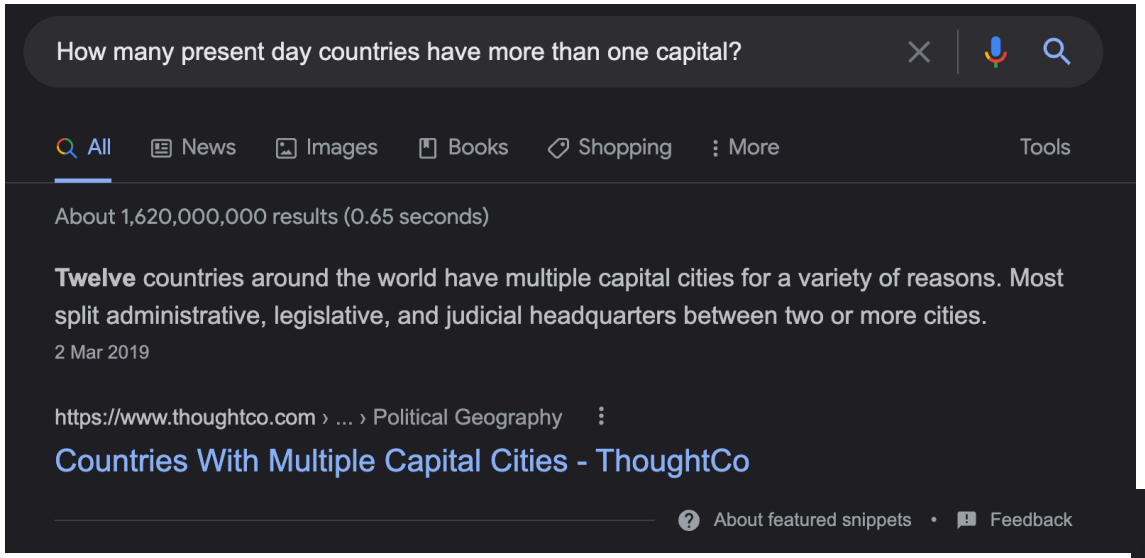
- The accuracy very much depends on
  - The domain
  - Open IE or ontology-based
  - What type of inference is required to infer the relation

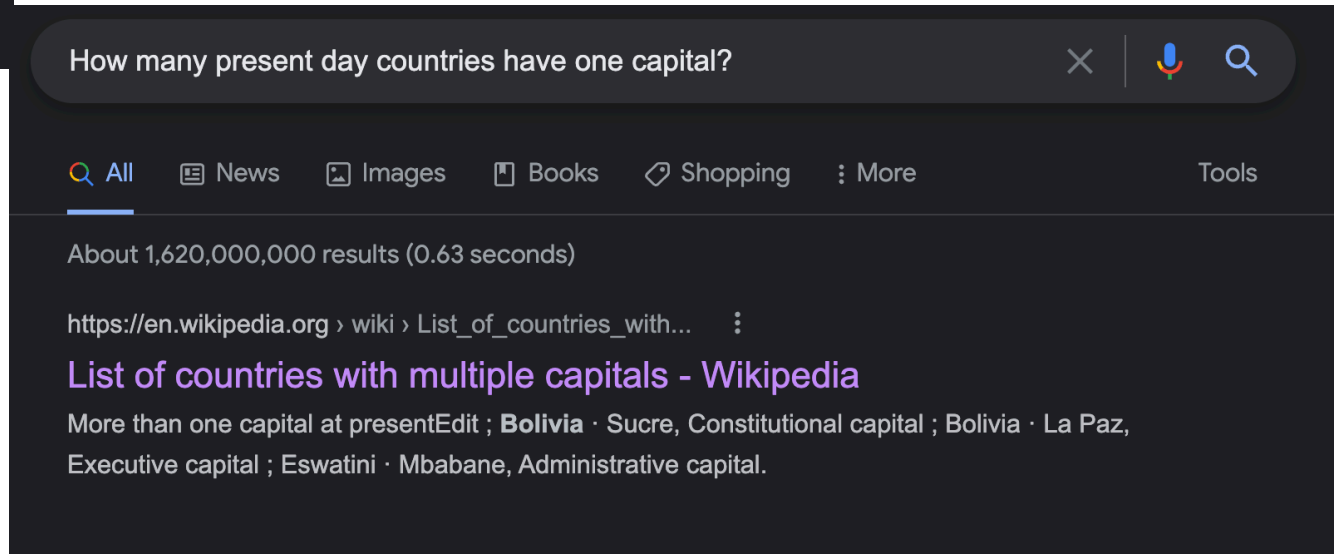
Question Answering (QA)

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



Question Answering (QA)





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 (generation, תספורת)
  - Morphemes can express many different meanings
    - התפעל can be reflexive התרחץ, הסתפר, התפטר but can also have other meanings התכתב ,התפלפל 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:
  - 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]

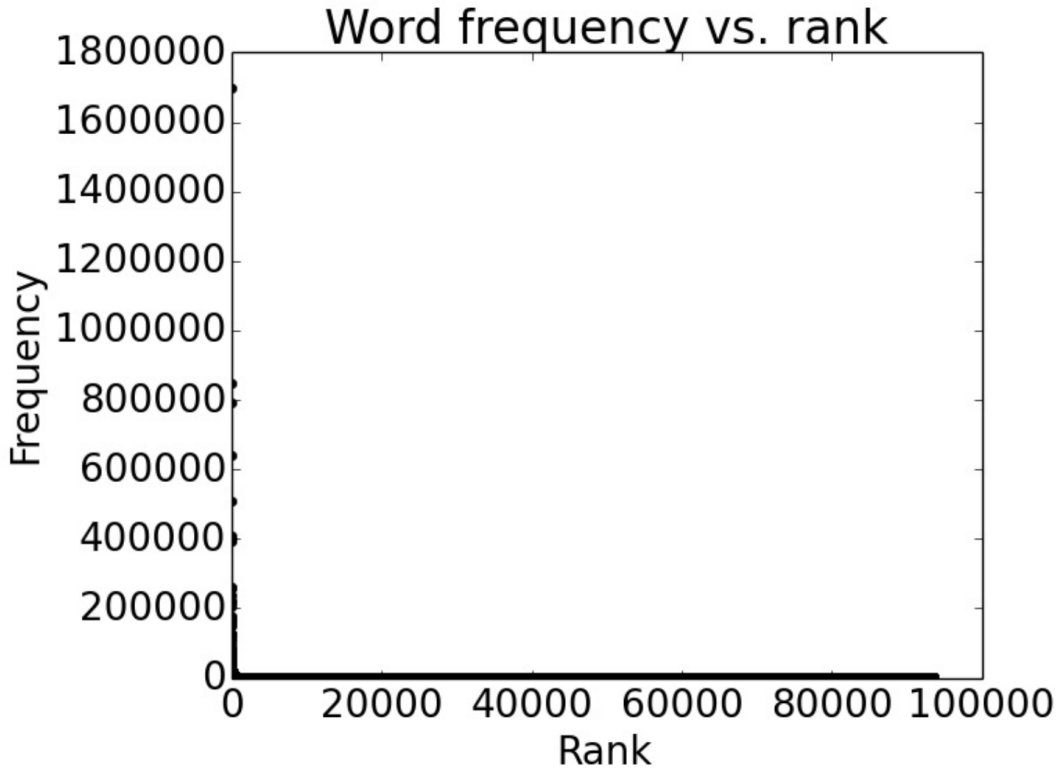
- But can't we overcome this if see enough examples of "Rock and Roll"?

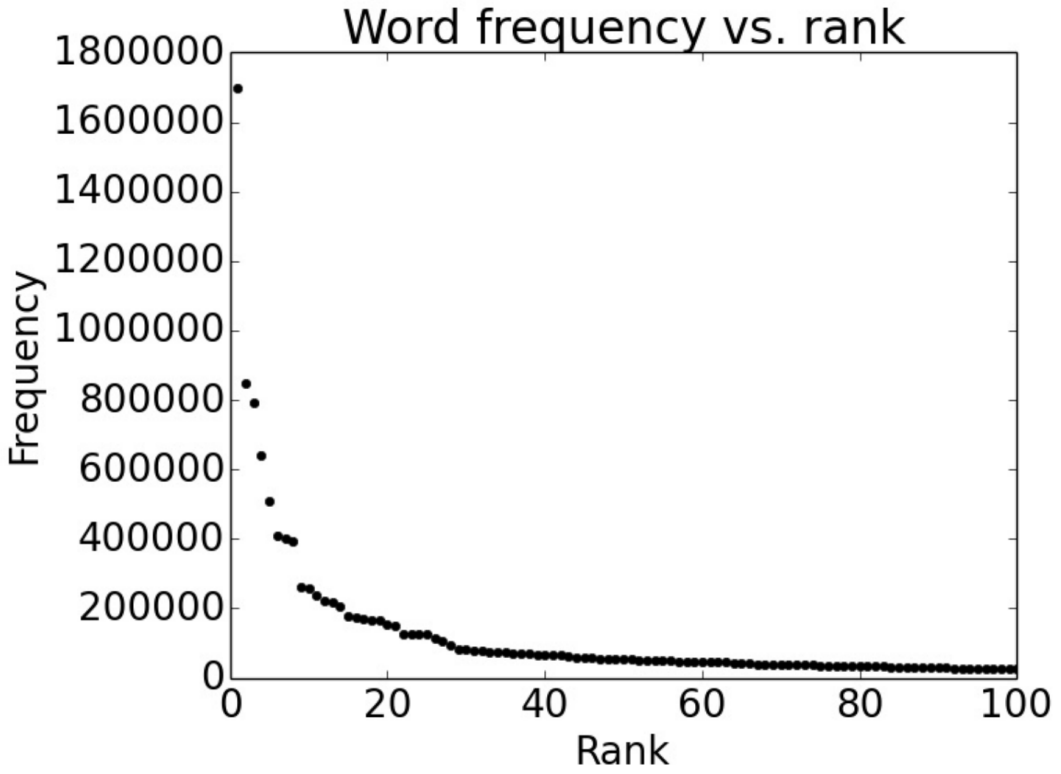
Key Challenges: Different Ways to Say the Same Thing

- He's passed on! This parrot is no more! He has ceased to be! He's expired and gone to meet its maker! He’s a stiff! Bereft of life, he rests in peace! If you hadn't nailed him to the perch he'd be pushing up the daisies! Its metabolic processes are now history! He’s off the twig! He’s kicked the bucket, He’s shuffled off its mortal coil, run down the curtain and joined the bleedin' choir invisible!! THIS IS AN EX-PARROT!!

- What does X do for a living? What is X’s profession? What does X do?
  - Knowing that they can mean the same thing is crucial for IE and QA

Sparsity


Word frequency vs. rank
Frequency
Rank
1800000
1600000
1400000
1200000
1000000
800000
600000
400000
200000
0
0
20000
40000
60000
80000
100000


Word frequency vs. rank
Frequency
Rank
1800000
1600000
1400000
1200000
1000000
800000
600000
400000
200000
0
0
20
40
60
80
100

Sparsity

- Zipf's law:

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

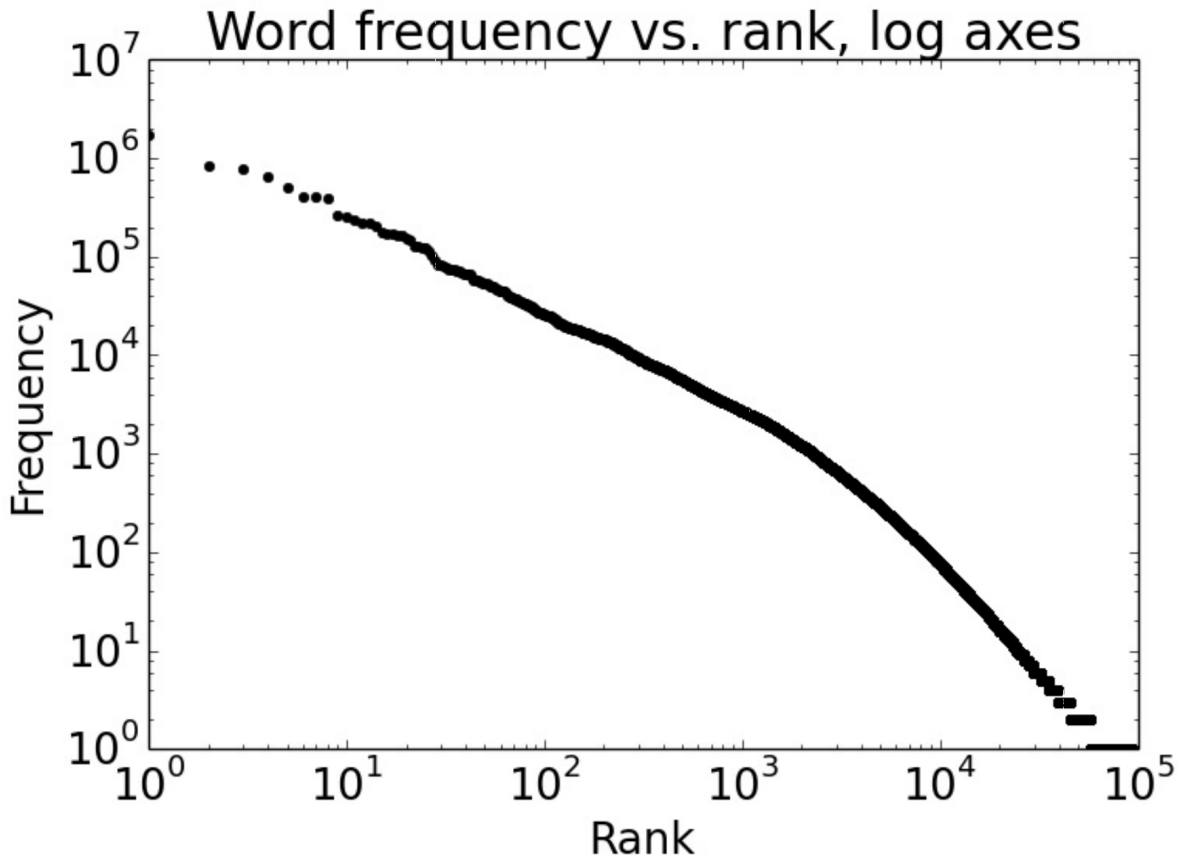
f = frequency of a word

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→V+ed, but many irregulars)
  - “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 “Mississipi River”

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

- *… anyone speaking a language possesses, implicitly, an enormous knowledge of the statistics of the language. Familiarity with the words, idioms, clichés and grammar enables him to fill in missing or incorrect letters in proof-reading, or to complete an unfinished phrase in conversation..* Claude Shannon, 1951

Layers of Linguistic Representation

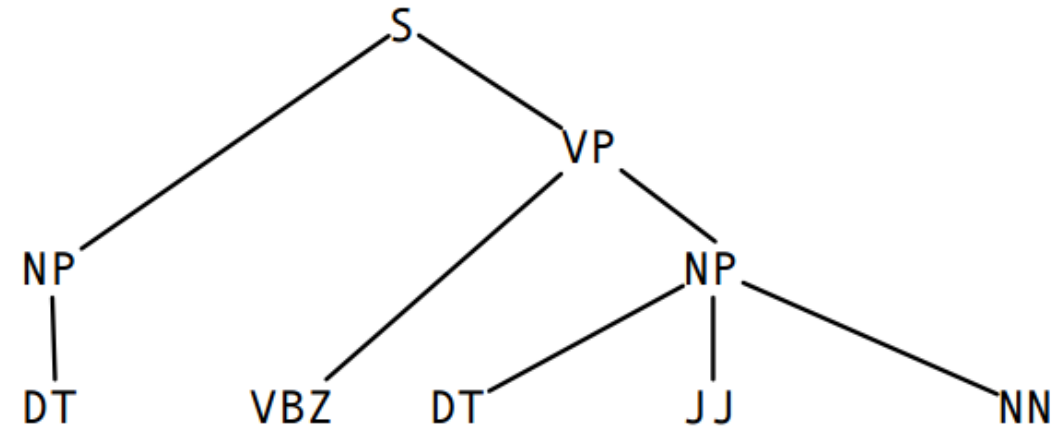
- Language consists of many levels of structure
- Humans fluently integrate all of these in producing and understanding language
- Some basic representations:

This	is	a	simple	sentence
det	*verb*	*det*	*adj*	*noun*
	be,3sg,pres			

having few parts

string of words satisfying the grammar of a language

Layers of Linguistic Representation



This is a simple sentence

det *verb* *det* *adj* *noun*

be,3sg,pres

But it is an instructive one

Contrast

having few parts *string of words satisfying the grammar of a language*

What will we cover?

Natural Language Processing

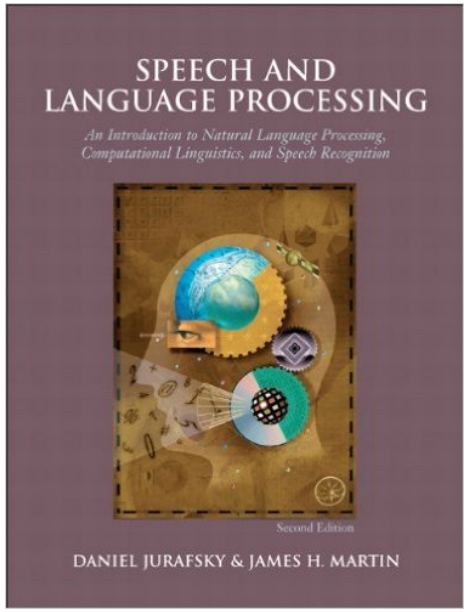
- Language models
- Document classification
- Part of speech tagging
- Syntax and Parsing: graph-based methods, transition-based methods
- Sentiment Analysis
- Sentence Semantics
- Information Extraction
- Machine Translation

Methods

- 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
- Recurrent Neural Networks
- Transformers and Pretraining

Useful Resources

- Speech and Language Processing, 2nd Edition, Jurafsky and Martin, 2008
  - Draft version of 3rd edition is on Dan Jurafsky's website
  - https://web.stanford.edu/~jurafsky/slp3/



- Introduction to Natural Language Processing, Jacob Eisenstein

- ACL Anthology (https://aclanthology.org/)

ACL Anthology

Useful Resources

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

- Google Scholar

Acknowledgments

- Some of the course Material was adapted from course materials by:
  - Sharon Goldwater and Nathan Schneider (University of Edinburgh course)
  - Ray Mooney (University of Texas)
  - Yoav Artzi (Cornell Tech)
  - Dan Jurafsky (Stanford)