

CS769 Advanced NLP

# Introduction to Natural Language Processing

Junjie Hu

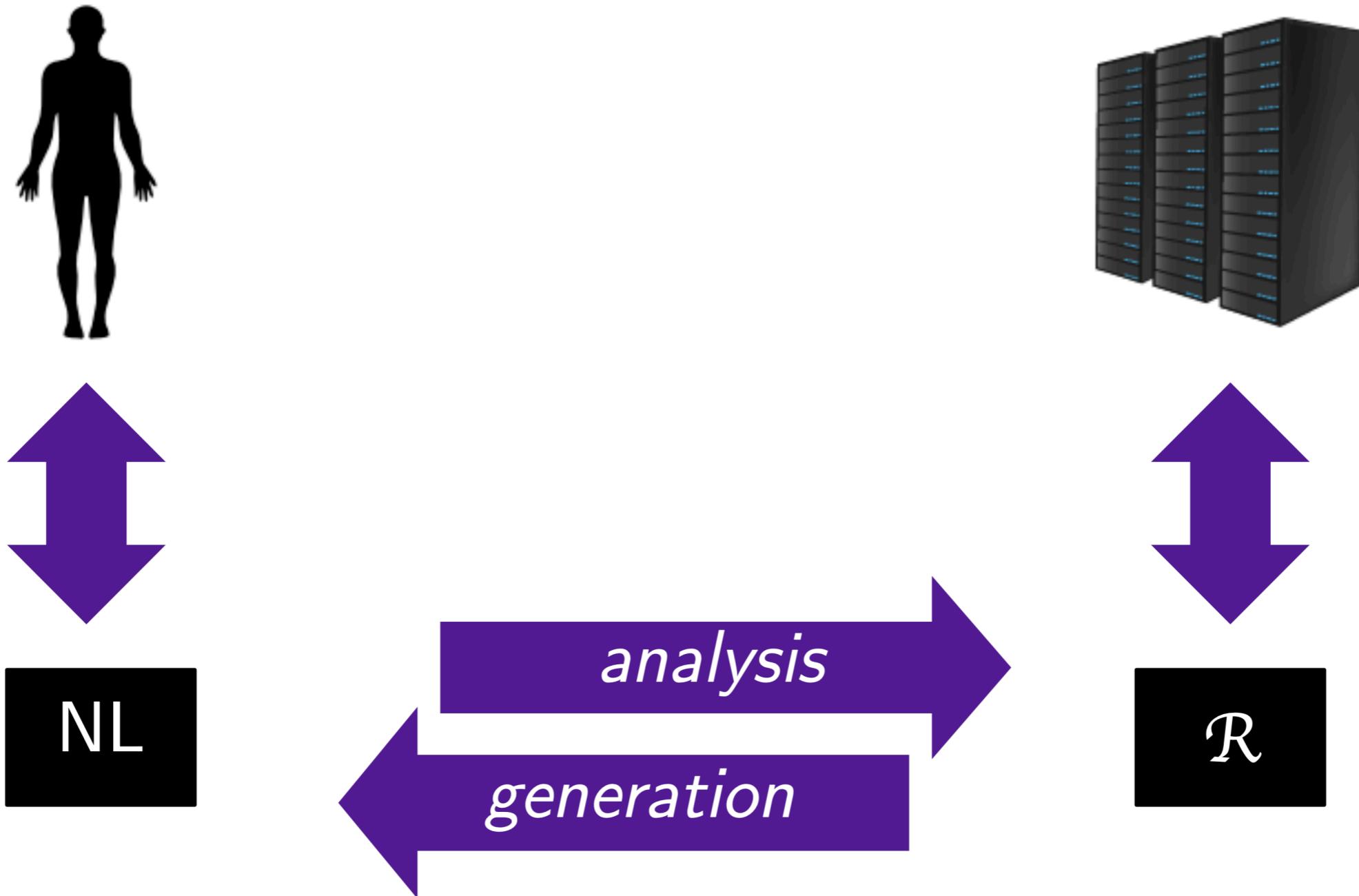


Slides adapted from Noah, Yulia, Graham  
<https://junjiehu.github.io/cs769-fall24/>

# What is NLP?

- $\text{NL} \in \{\text{Chinese, English, Spanish, Hindi, ...}\}$
- $\mathcal{R}$ : *intermediate meaning representations*
- Automation of:
  - **Analysis** or Interpretation of what a text means ( $\text{NL} \rightarrow \mathcal{R}$ )
  - **Generation** of fluent, meaningful text
  - **Acquisition** of these capabilities from knowledge and data

# What is NLP?

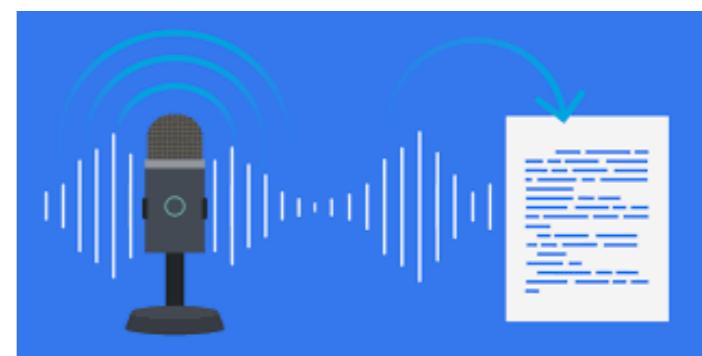
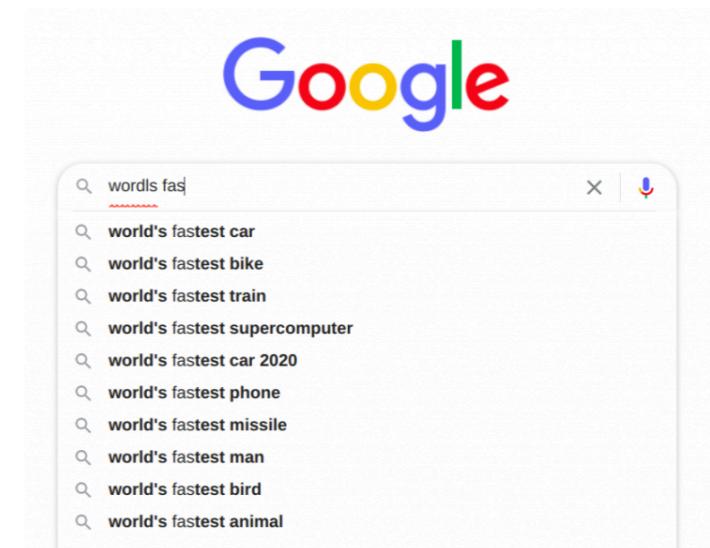
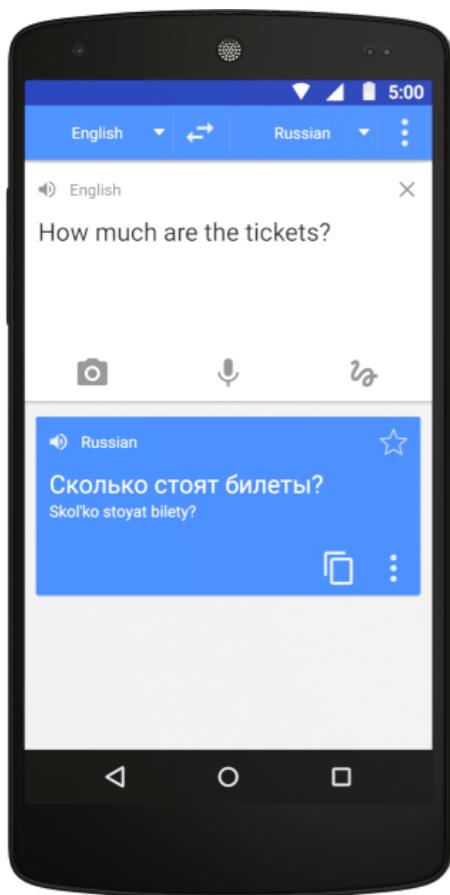


# Why NLP?

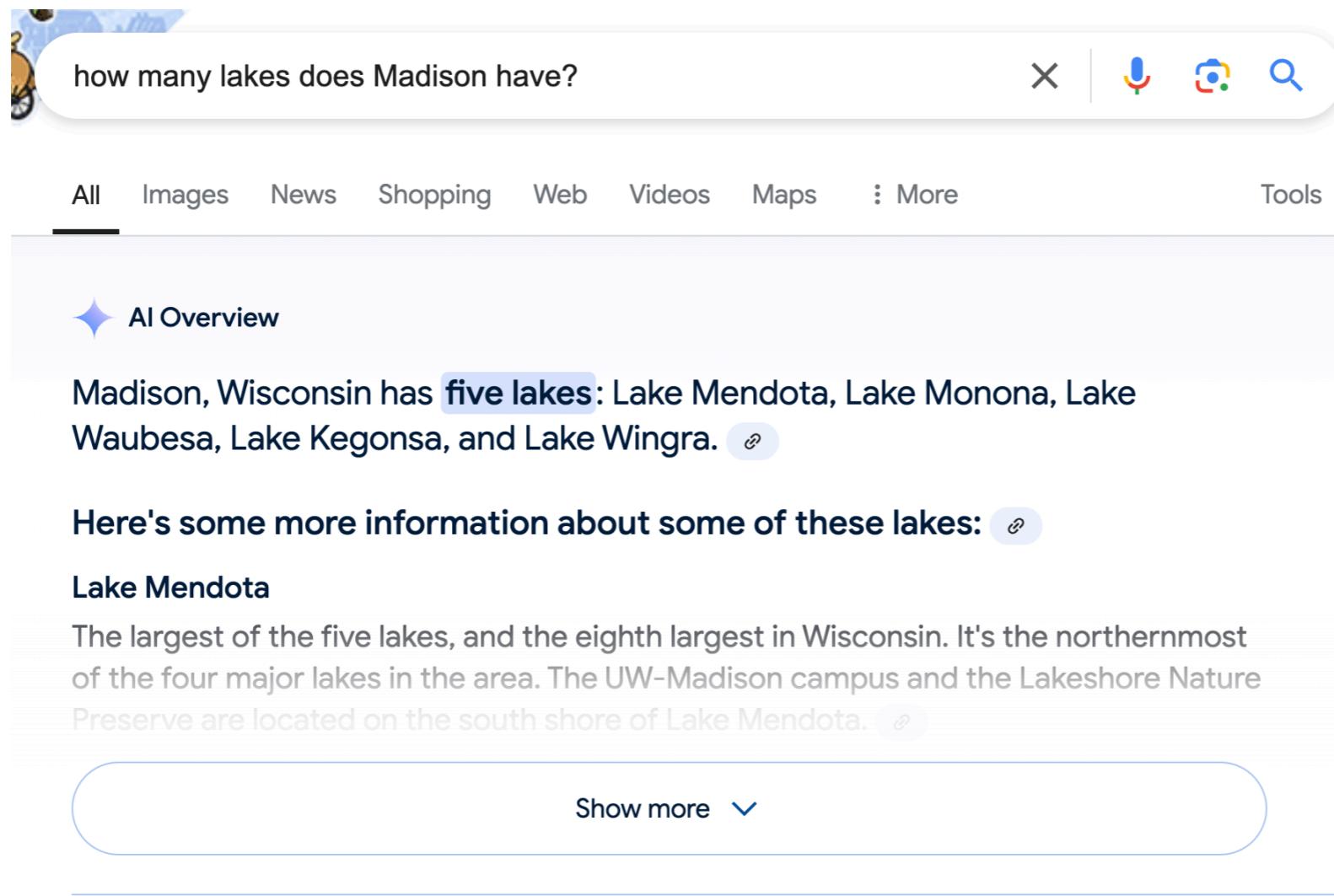
- Technology to handle human language (usually text) using computers
- Aid **human-human communication** (e.g., machine translation)
- Aid **human-machine communication** (e.g., question answering, dialog systems)
- **Analyze/generate language** (e.g., syntactic analysis, text classification, entity/relation recognition/linking)

# Language Technologies

- We now use NLP several times a day, sometimes without knowing it!



# NLP can Answer our Questions



how many lakes does Madison have?

All Images News Shopping Web Videos Maps More Tools

AI Overview

Madison, Wisconsin has **five lakes**: Lake Mendota, Lake Monona, Lake Waubesa, Lake Kegonsa, and Lake Wingra. [🔗](#)

Here's some more information about some of these lakes: [🔗](#)

**Lake Mendota**

The largest of the five lakes, and the eighth largest in Wisconsin. It's the northernmost of the four major lakes in the area. The UW-Madison campus and the Lakeshore Nature Preserve are located on the south shore of Lake Mendota. [🔗](#)

Show more ▾

## five lakes

Madison's Lakes - There are **five** lakes that make up Madison's waterways.



Retrieved Sept 4, 2024

# NLP can Answer our Questions

- What does “divergent” mean?
- What year was Abraham Lincoln born?
- How many states were in the United States that year?
- How much Chinese silk was exported to England in the end of the 18th century?
- What do scientists think about the ethics of human cloning?



IBM Watson won Jeopardy!  
Champions in early 2011

# NLP can Translate Text

The screenshot shows the Google Translate interface. At the top, there are two tabs: "Text" (selected) and "Documents". Below the tabs, language detection and selection are shown: "DETECT LANGUAGE" (English), "CHINESE", and "CHINESE (SIMPLIFIED)" (selected). There is also a "TRANSLATE" button with a double arrow icon between the source and target language dropdowns. The main area displays a paragraph about Madison, Wisconsin, in English on the left and its Chinese translation on the right. The English text reads:

Madison is the capital of the U.S. state of Wisconsin and the seat of Dane County. As of the 2020 census the population was 269,840 which made it the second-largest city in Wisconsin by population, after Milwaukee, and the 80th-largest in the United States. The city forms the core of the Madison Metropolitan Area which includes Dane County and neighboring Iowa, Green, and Columbia counties for a population of 680,796. Madison is named for American Founding Father and President James Madison.

The Chinese translation is:

麦迪逊是美国威斯康星州的首府，也是戴恩县的所在地。截至 2020 年人口普查，人口为 269,840，使其成为威斯康星州人口第二大城市，仅次于密尔沃基，在美国排名第 80。该市是麦迪逊都会区的核心，包括戴恩县和邻近的爱荷华县、格林县和哥伦比亚县，人口达 680,796 人。麦迪逊以美国国父和总统詹姆斯麦迪逊的名字命名。

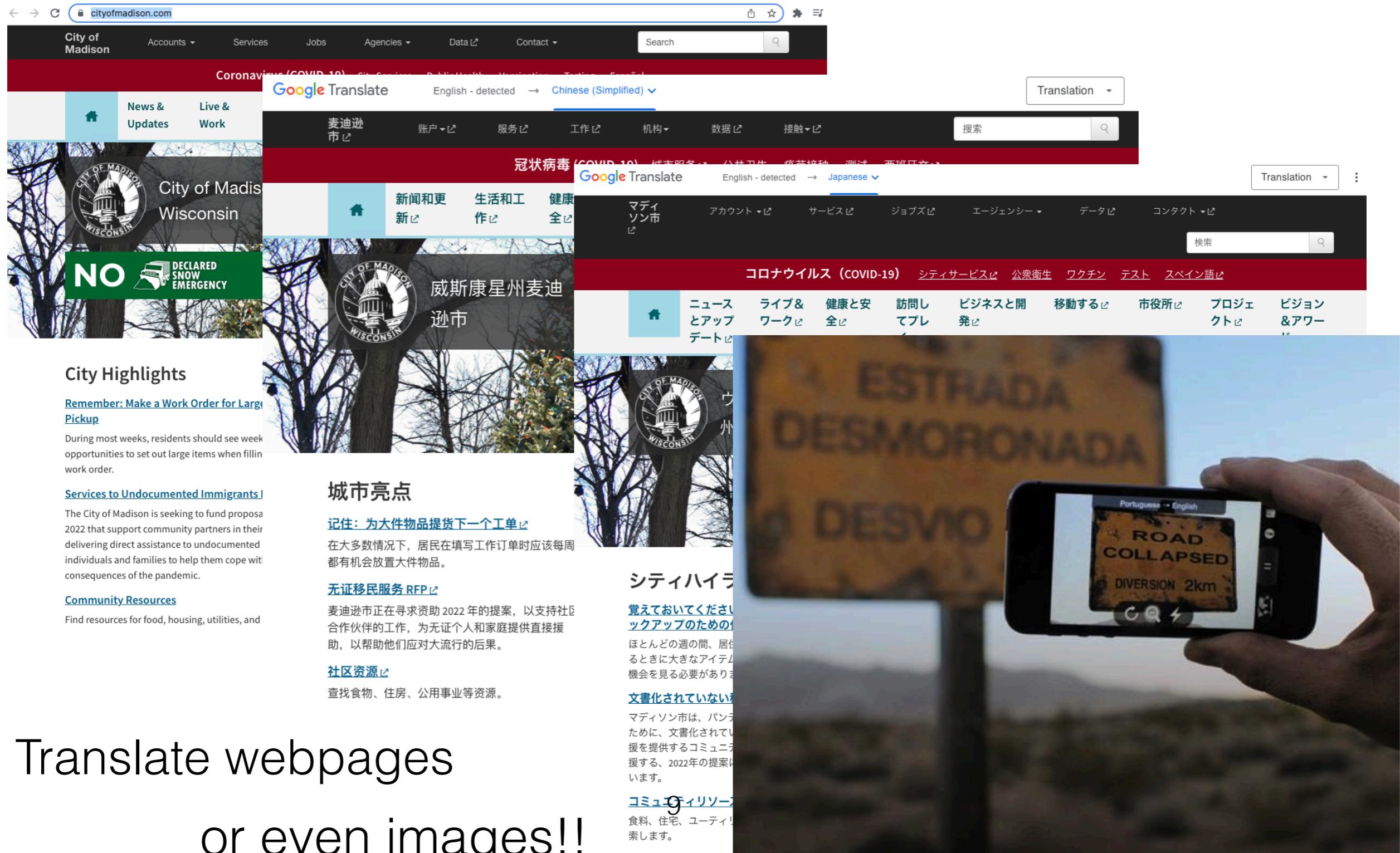
Below the text, there is a phonetic transcription in Pinyin: Mǎidí xùn shì měiguó wēisīkāngxīng zhōu de shǒufǔ, yěshì dài ēn xiàn de suǒzàidì. Jiézhì 2020 nián rénkǒu pǔchá, rénkǒu wèi 269,840, shǐ qí chéngwéi wēisīkāngxīng zhōu

[Show more](#)

At the bottom of the interface, there are various icons for audio playback, document sharing, and feedback submission. The status bar at the bottom right says "Send feedback".

English Wikipedia, translated by Google Jan. 25, 2022

# NLP can Translate Text

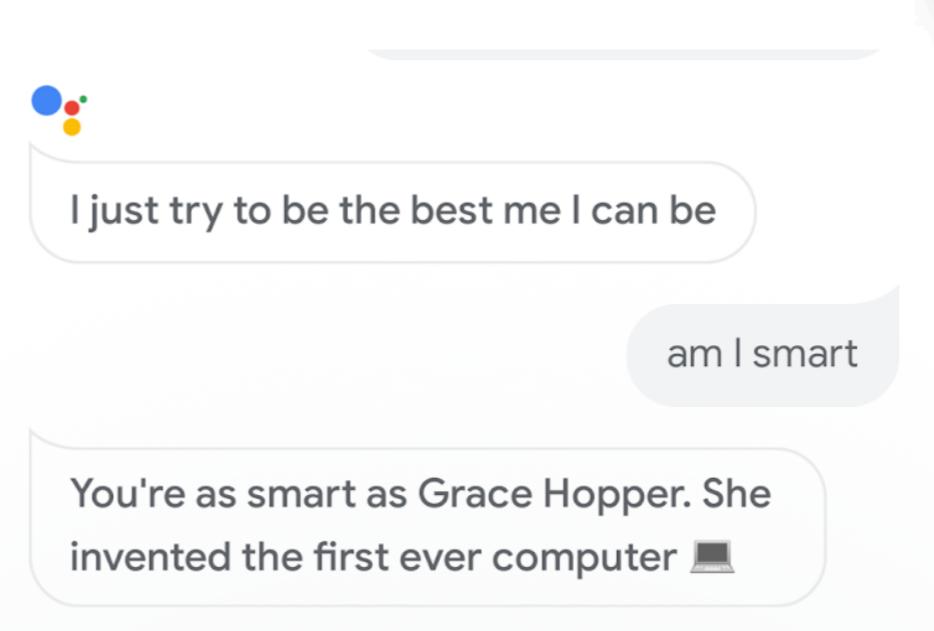
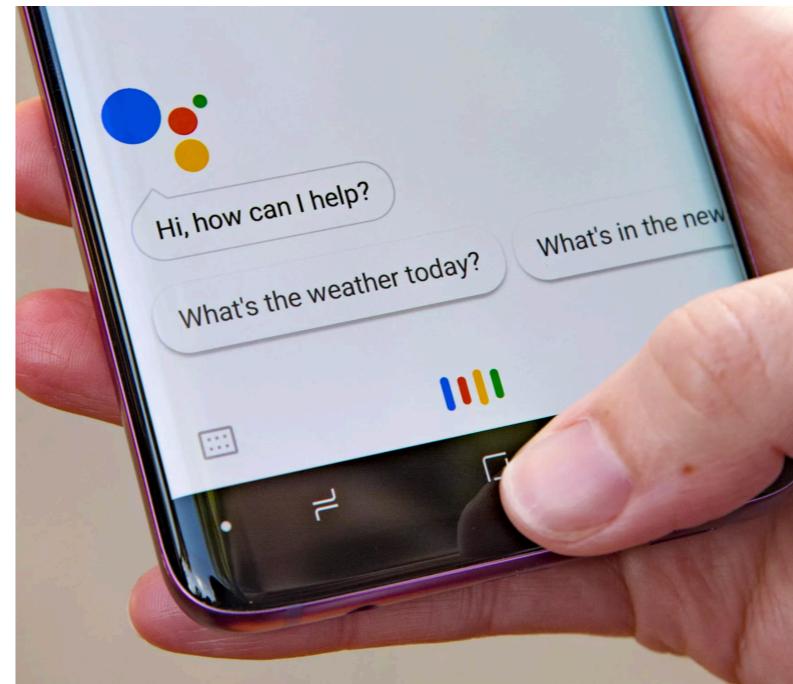


Translate webpages  
or even images!!

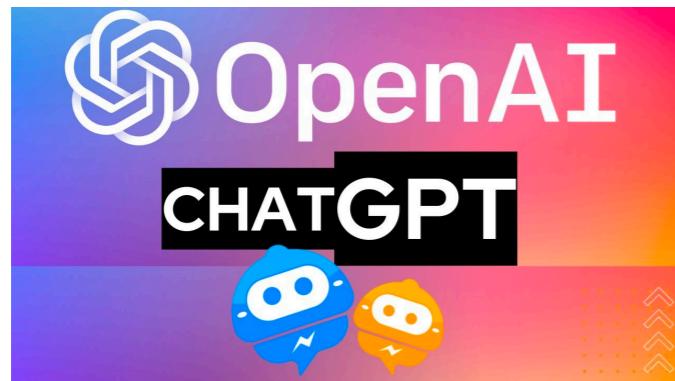
# NLP can Chat with You

Conversational agents:

- Speech recognition
- Language analysis
- Dialogue processing
- Information retrieval
- Text to speech



# LLMs can be prompted with language instructions



< 3/3 >



## Job Posting: Post-Doctoral Research Scientist Position at UW-Madison

The Natural Language Processing (NLP) Research Group at the University of Wisconsin-Madison invites applications for a Post-Doctoral Research Scientist position. This position is part of an exciting NSF-funded project focusing on the development of preference alignment algorithms for large language models.

### Position Overview

We are seeking a talented and highly motivated post-doctoral research scientist to contribute to cutting-edge research at the intersection of NLP, machine learning, and preference modeling. The successful candidate will work on designing, implementing, and evaluating algorithms that enable better preference alignment in large language models, helping to advance the state of the art in this rapidly growing field.

### Key Responsibilities

- Conduct original research in preference alignment algorithms for large language models.
- Develop novel methodologies to enhance alignment between model outputs and human preferences.
- Collaborate with a dynamic research team to publish in top-tier conferences and journals.
- Contribute to open-source projects and share findings with the broader NLP community.

### Qualifications

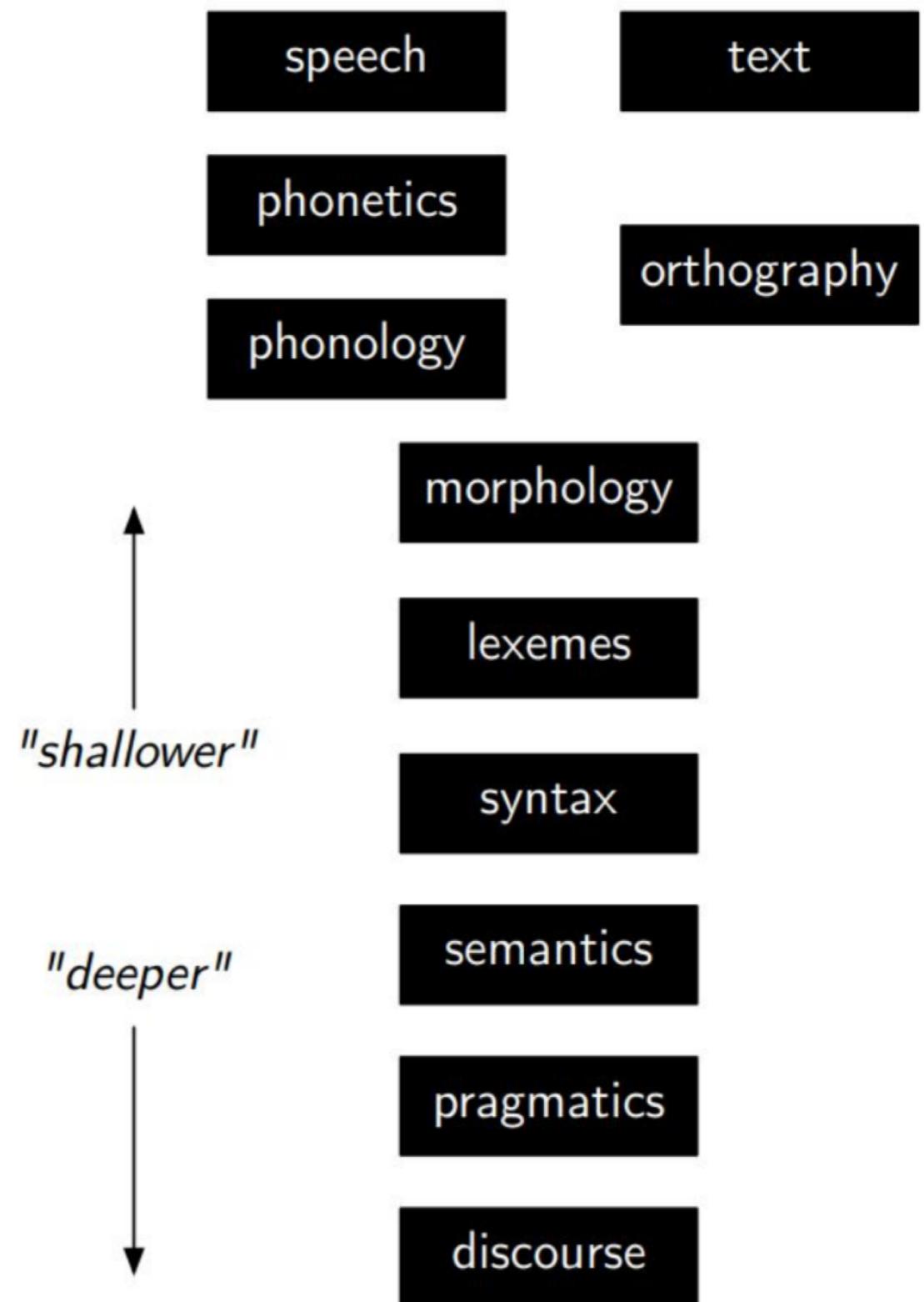
- PhD in Computer Science, Machine Learning, Artificial Intelligence, or a related field.

# NLP

- **Applications**
  - Machine translation
  - Information retrieval
  - Question answering
  - Dialogue systems
  - Information extraction
  - Summarization
  - Sentiment analysis
  - ...
- **Core technologies**
  - Language modeling
  - Part-of-speech tagging
  - Syntactic parsing
  - Named entity recognition
  - Coreference resolution
  - Word sense disambiguation
  - Semantic role labelling
  - ...

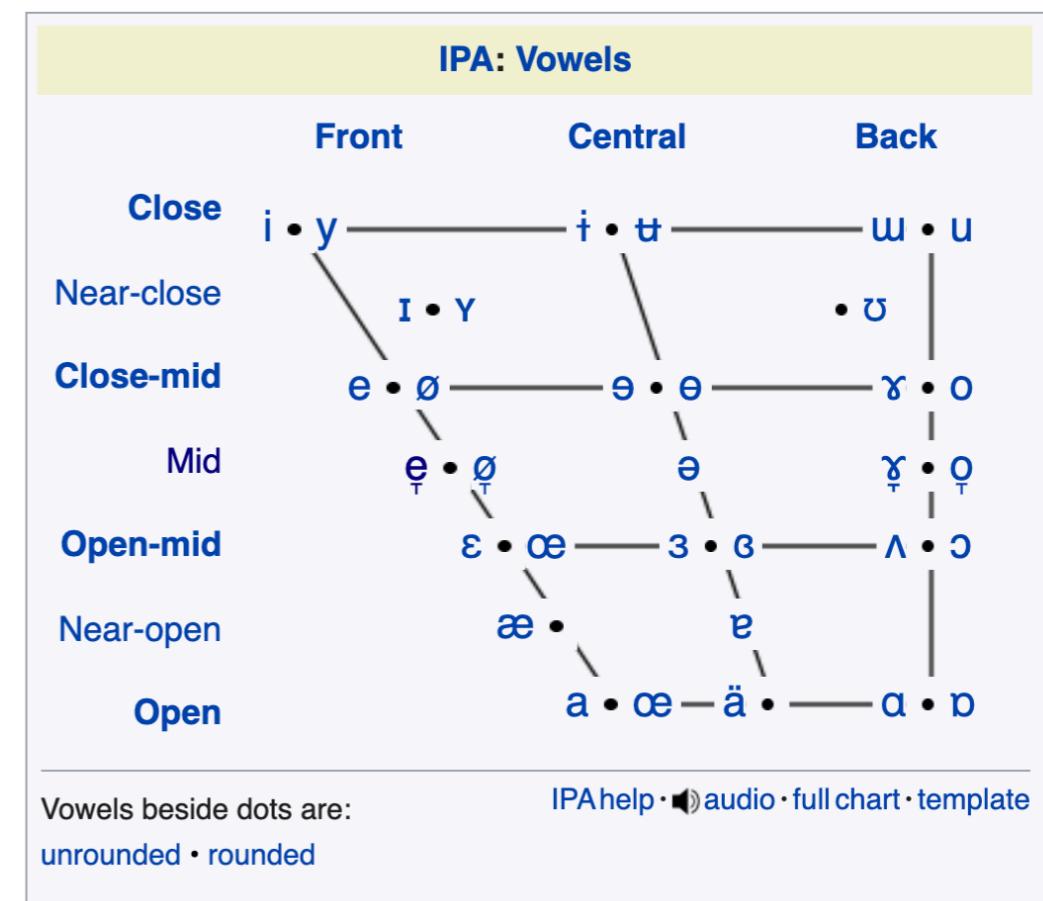
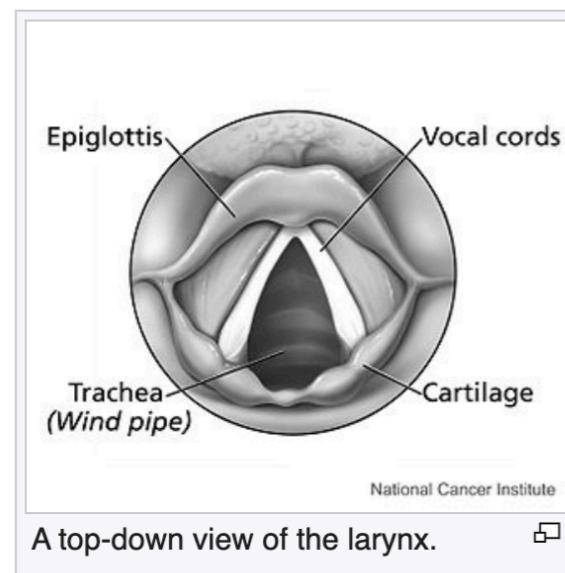
# Levels of Linguistic Knowledge

- What does an NLP system need to “know” a language?



# Phonetics, Phonology

- Study how humans produce and perceive **sounds**, or in the case of sign languages, the equivalent aspects of **sign**



# Orthography (Writing Systems)

- Thai script:
  - ลูกศิษย์วัดกระทิงยังชี้อปิดถนนทางขึ้นไปนมัสการพระบาทเจ้าคิชณกุฎ หวิดปะทะ กับเจ้าถินที่ออกมาเผชิญหน้า เพราะเดือดร้อนสัญจรไม่ได้ ผจ.เร่งทุกฝ่ายเจรจา ก่อนที่ชื่อเสียงของจังหวัดจะเสียหายไปมากกว่านี้ พร้อมเสนอหยุดจัดงาน 15 วัน....
- Latin script:
  - The Latin script, also known as Roman script, is an alphabetic writing system based on the letters of the classical Latin alphabet.
- Arabic script:
  - لم تعرف منظمة الأمم المتحدة باللغة العربية رسمياً إلا في 18 ديسمبر عام 1973، بعد محاولات مضنية من قادة الدول العربية للاعتراف باللغة العربية داخل المنظمة الأممية الكبيرة منذ تأسيسها عام 1945 وحتى تاريخ الاعتراف.
  - طيلة السنوات التي لم تكن الأمم المتحدة اعترفت باللغة العربية رسمياً، كان رؤساء الدول العربية يتحدثون اللغة العربية مع حضور مترجم، وكان أول رئيس يقوم بإلقاء خطاب سياسي قبل قرار الاعتراف هو رئيس جمهورية مصر العربية جمال عبد الناصر.

.. المزيد على دنيا الوطن

# Morphology (Assembly of Words)

- Study of how words are formed: such as stems, root words, prefixes, suffixes
  - [Turkish]: uygarla,stıramadıklarımızdanmı,ssınızcasına ← agglutinative language
  - [English]: “(behaving) as if you are among those whom we could not civilize”
  - [English]: unfriend → un + friend , Obamacare → Obama + care

WORDS	This is a simple sentence
MORPHOLOGY	be 3sg present

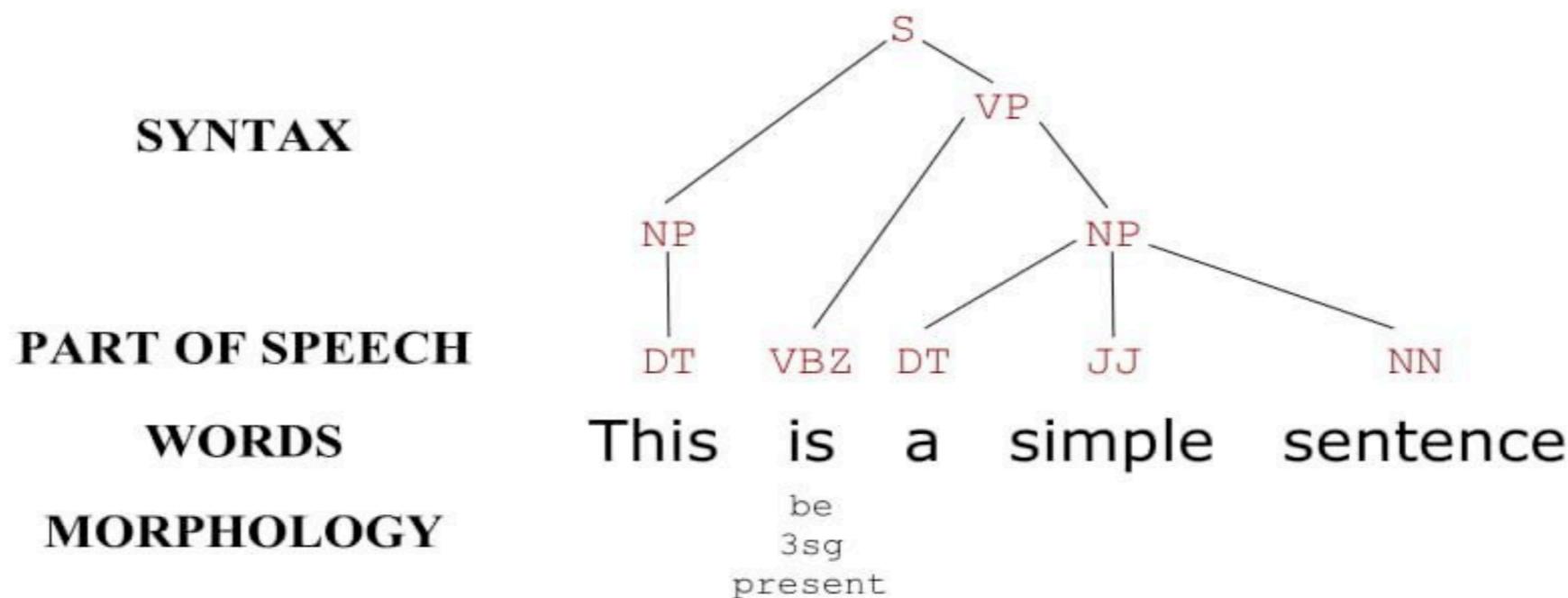
# Part-of-Speech

- Predict which category a word is assigned to in accordance with its syntactic functions.

<b>PART OF SPEECH</b>	DT	VBZ	DT	JJ	NN
<b>WORDS</b>	This	is	a	simple	sentence
<b>MORPHOLOGY</b>		be			
		3sg			
		present			

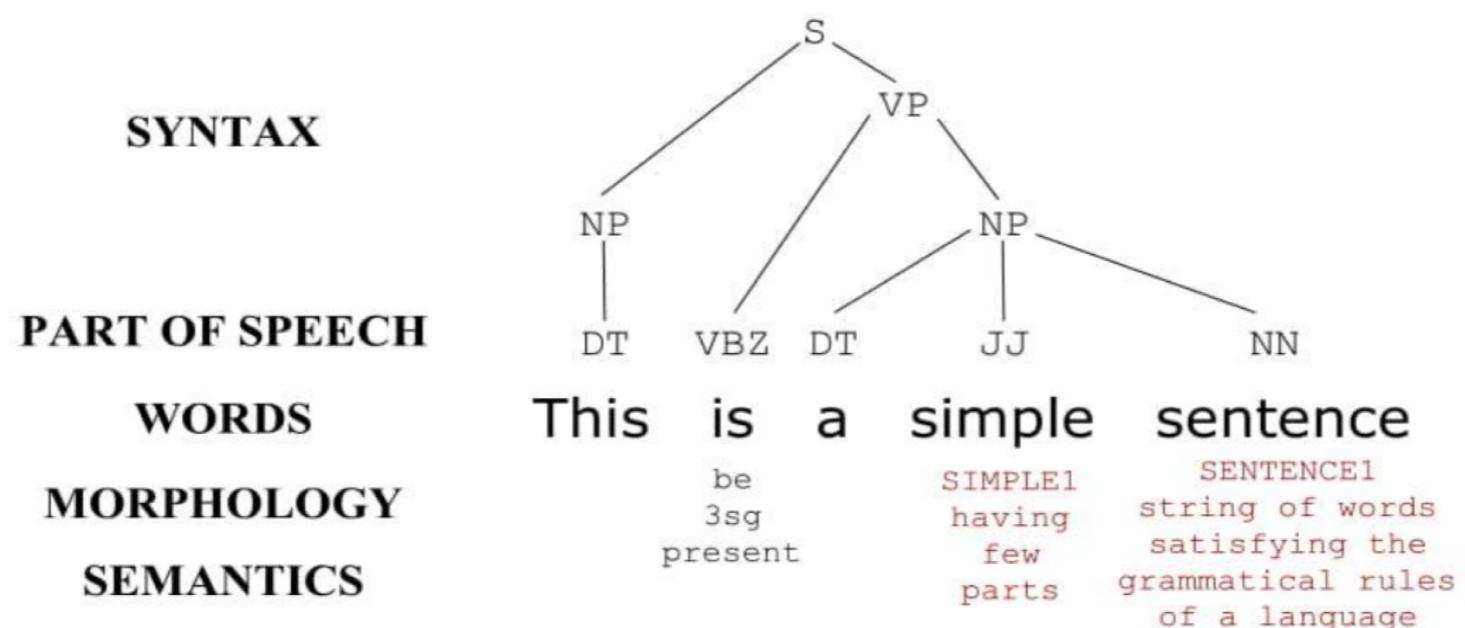
# Syntax

- Study of how words and **morphemes** combine to form larger units such as **phrases** and **sentences**.
  - Constituency Grammars
  - Dependency Grammars



# Semantics

- Study meaning of words, phrases, sentences, or larger units (w/ discourse)
  - Named entity recognition
  - Word sense disambiguation
  - Semantic role labeling



# Discourse

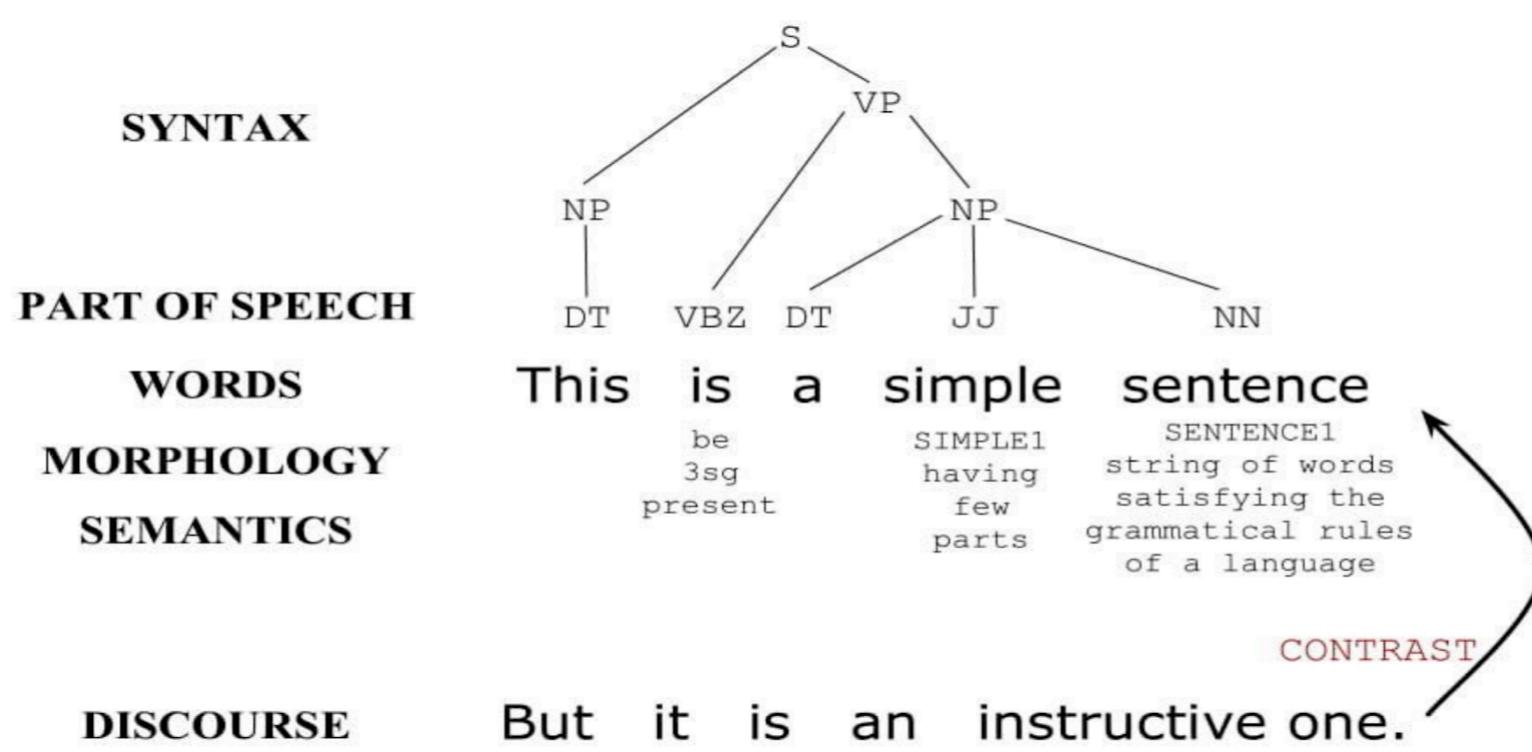
- Analysis of language “beyond the sentence”

<> analysis of sounds (phonetics)

<> analysis of words (morphology)

<> analysis of meaning (semantics)

<> analysis of word order (syntax)



Where are we now for NLP  
research?

# NLP cannot Answer our Questions

who won the 2021 pittsburgh mayor democratic primary? X

All News Images Videos Web Shopping Forums More Tools

**Wikipedia**  
[https://en.wikipedia.org › wiki › 2021\\_Pittsburgh\\_mayo... :](https://en.wikipedia.org/wiki/2021_Pittsburgh_mayor_election)

**2021 Pittsburgh mayoral election**

The **2021 Pittsburgh mayoral election** took place on November 2, 2021. The **primary election** was held on May 18, 2021. The **Democratic** nominee, State ...

The **2021 Pittsburgh mayoral election** is scheduled to take place on November 2, 2021. The **primary election** was held on May 18, 2021. Incumbent Democratic Mayor **Bill Peduto** ran for re-election to a third term in office, but **lost renomination** to state representative **Ed Gainey**.<sup>[1]</sup> Four Democrats and no Republicans filed to appear on their respective primary

Retrieved Sept. 5 2024

# NLP cannot Answer our Questions

who invented neural machine translation

All News Images Videos Shopping More Tools

About 725,000 results (0.63 seconds)

English - detected

French

who invented  
neural machine

qui a inventé la  
machine neuronale

Open in Google Translate • Feedback

Retrieved Aug. 29, 2021

# NLP cannot Translate Text

“၃၇၊ ၃၈ မာရိလေးမဖို့ခဲ့ ၈၄ မိန္ဒီးလမ္မားမခကီးပေါ့၊ စုတုန်းပဲ  
ရှိဝေသေးတယူဝေပါဝေနာ့။ ထြက်ဗျာအတွက်၊ အဲဒါကို သူတို့၏အာ  
ကေန ခာကီတင့် သတင့်းရလာတယ့် မသိဘူး၊ ခုက္ခာင်း ဝေရာက္ခာလာမ  
ပါးဝေတဲ့ အဟန္တမုံးဖက် ဝါ့ဝေရာက္ခာပို့ခြင်းတေပါ့။ အတိအက်  
တဲ့ ကြောန်ဝေတာ့တို့၏လဲ မသိရေသားဘူး။ ၄ ဝေယာက္ခာ့တယ့်  
လို့၏လဲ ဝေချုဟတယ့်။ ၆ ဝေယာက္ခာ့တယ့်လို့၏လဲ ဝေချုဟတယ့်။ ဘ  
ယ့်ဝေလာက္ခာ့တယ့်လဲဆိုတာ ခုအခို့ အတိအက် မသိရေသားဘူး။ တ  
ခို့ပေါ်တို့ ဝေရာက္ခာဝေတာင့် မေရာက္ခာကေသားဘူး။ စစ်ဝေက္ခာနှုံဖက်  
အဟန္တမုံးဖက် ဖု့ခြင်းလို့ကဲဝေတဲ့ လက္ခာဝေတဲ့ အမွတ့ ၆ ထဲမှာ ဖ  
မူးခံထားရတယ့်၌ အဲဝေလာက္ခာ့ သိရေသားတယ့်။”



"37," he said. 38. 84 Main Road of Mandalay. I'm still collecting. I don't know where they got the information in advance. It arrived immediately and was violently suppressed. We do not know exactly. He said four people were involved. He also said that six people were involved. It is unknown at this time what he will do after leaving the post. Some have not even arrived. He is currently being held in No. 6 after a violent crackdown by the military junta.

Front page news from Voice of America Burmese, translated by Google Jun 25., 2021

# In this Class, we Ask:

- Why do current state-of-the-art NLP systems **work uncannily well** sometimes?
- Why do current state-of-the-art NLP systems still **fail**?
- How can we
  - **create systems for various tasks,**
  - **identify their strengths and weaknesses,**
  - **make appropriate improvements,**
  - and **achieve whatever we want to do with NLP?**

# Why NLP is Hard?

- Ambiguity
- Scale
- Sparsity
- Variation
- Expressivity
- Unmodeled variables
- Unknown representations  $R$

# Ambiguity

- Ambiguity at multiple levels:

- Words with multiple meanings: *bank* (finance or river?)
- Domain-specific meanings: *latex*
- Part-of-speech: *chair* (noun or verb?)
- Multiple meanings: *I made her duck.* →
  - I cooked waterfowl for her
  - I cooked waterfowl belonging to her
  - I created the (plaster?) duck she owns
  - I magically turned her into a duck

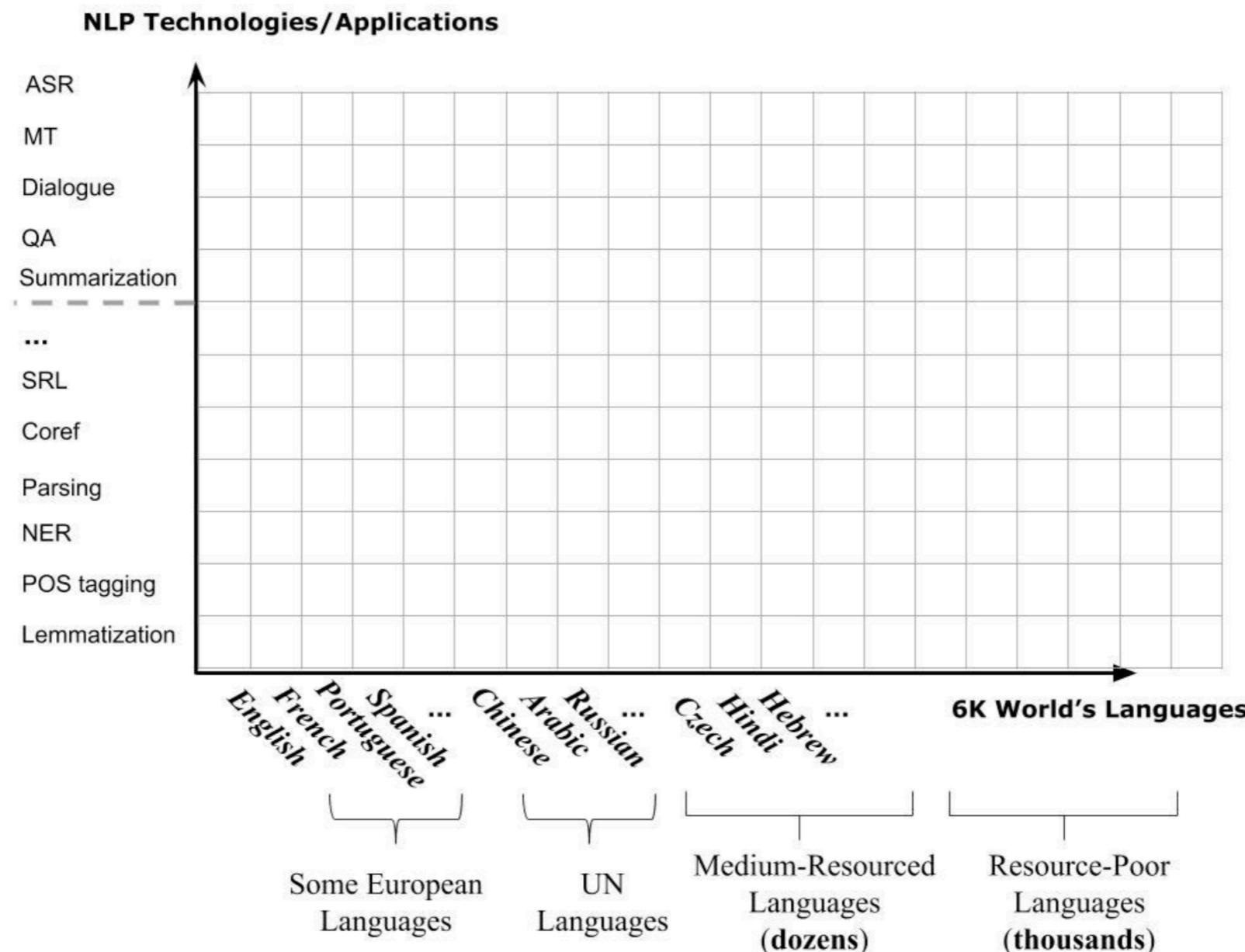


# More Challenges of “Words”

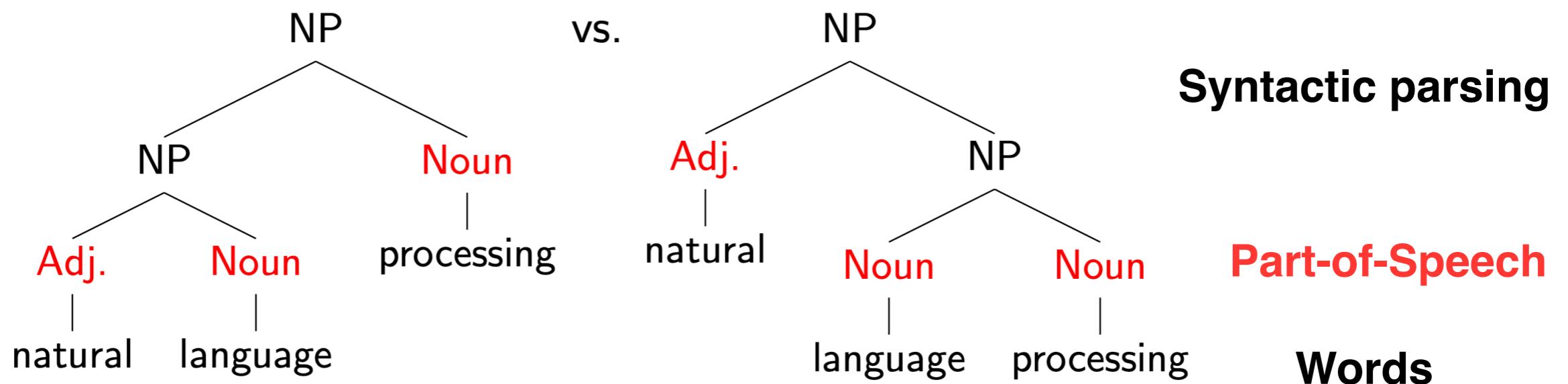
- Segmenting text into words (e.g., Thai example)
- Morphological variation (e.g., Turkish example)
- Multiword expressions: *take out*, *make up*
- New words (e.g., *covid*) and changing meanings (e.g., *Bachelor*: a young knight → an academic degree)

# Ambiguity + Scale

- Scale up to different **languages & tasks**.

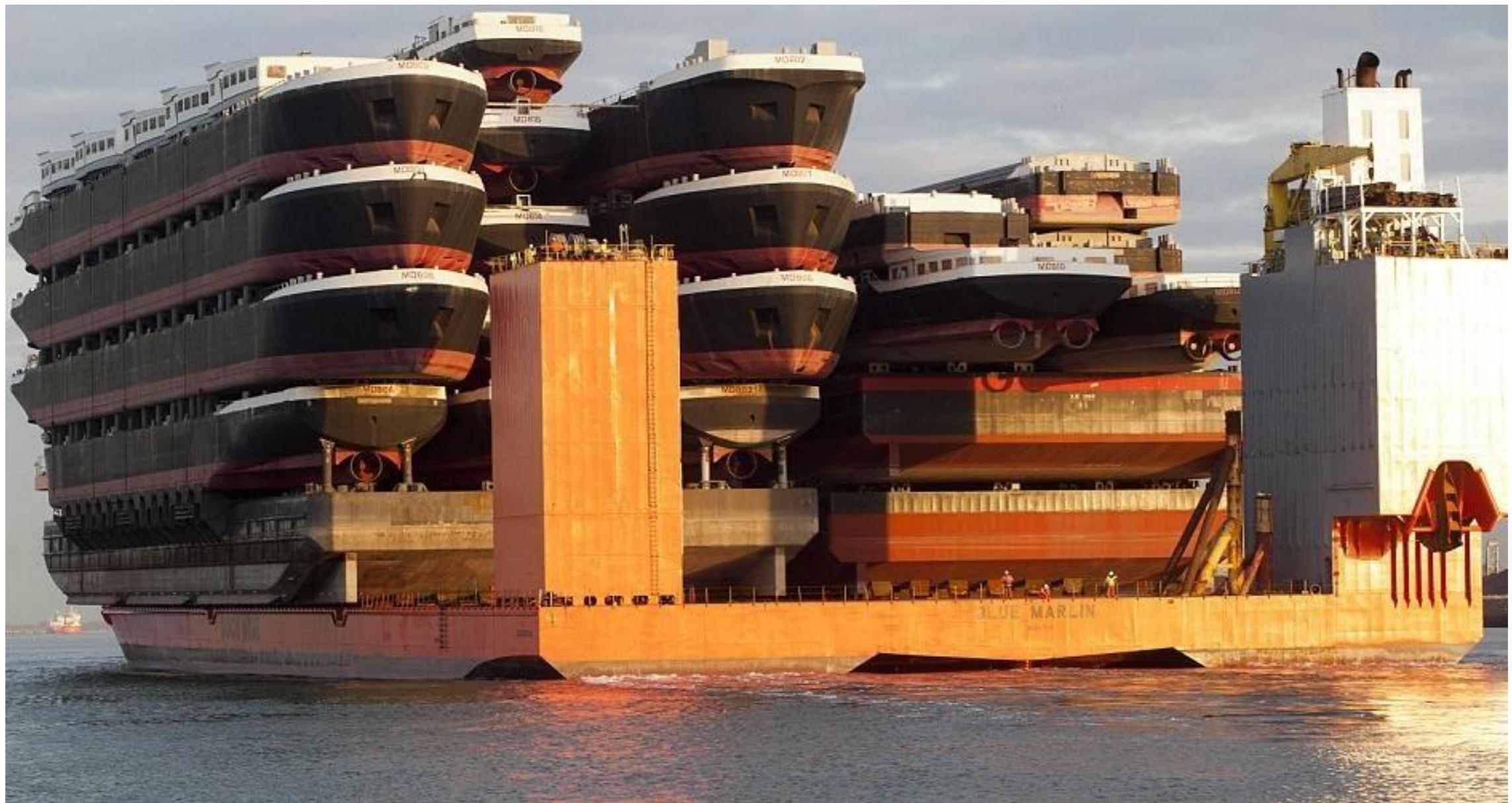


# Syntax Ambiguity



# Morphology + Syntax

- A ship-shipping ship, shipping shipping-ships



# Syntax + Semantic

We saw the woman with the telescope wrapped in paper.

- Who has the telescope?
- Who or what is wrapped in paper?
- An event of perception, or an assault?

# Semantic Ambiguity

- Every fifteen minutes a woman in this country gives birth.

# Semantic Ambiguity

- Every fifteen minutes a woman in this country gives birth. Our job is to find this woman, and stop her!

- Groucho Marx

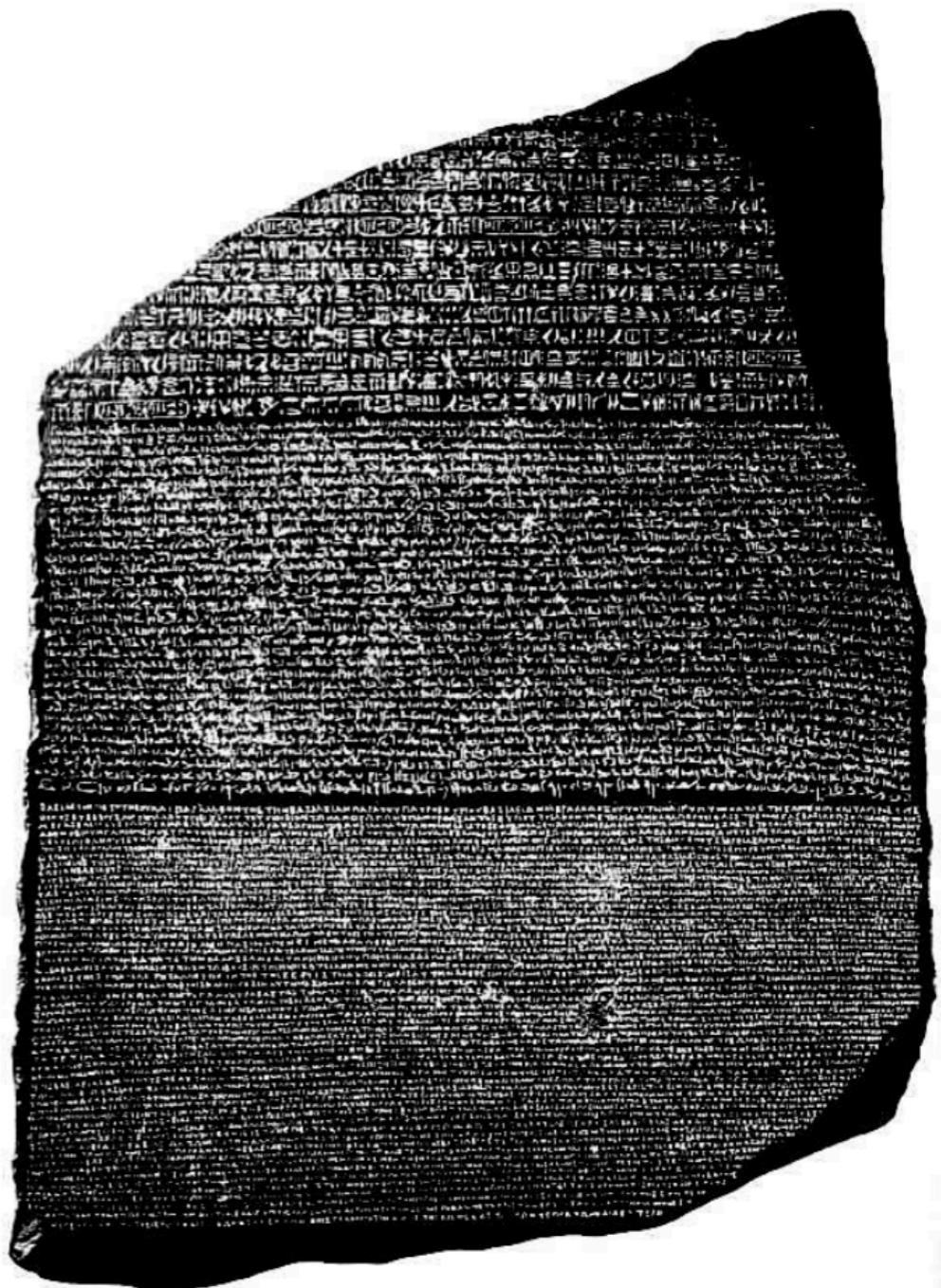


# Dealing with Ambiguity

- How can we model ambiguity and choose the correct analysis in context?
  - Non-probabilistic methods (Finite-state machines for morphology, CKY parsers for syntax) return *all possible analyses*.
  - Probabilistic models (HMMs for POS tagging, PCFGs for syntax) and algorithms (Viterbi, probabilistic CKY) return *the best possible analysis*
- But the “best” analysis is only good if our probabilities are accurate. Where do they come from?

# Corpora

- **A corpus is a collection of text**
  - Often annotated in some way
  - Sometimes just lots of text
- **Examples**
  - Penn Treebank: 1M words of parsed WSJ
  - Canadian Hansards: 10M+ words of aligned French/English sentences
  - Web: billions of words
  - Amazon reviews



# Corpus-based Methods

- Give us statistical information by counting
  - Example: Syntax parsing

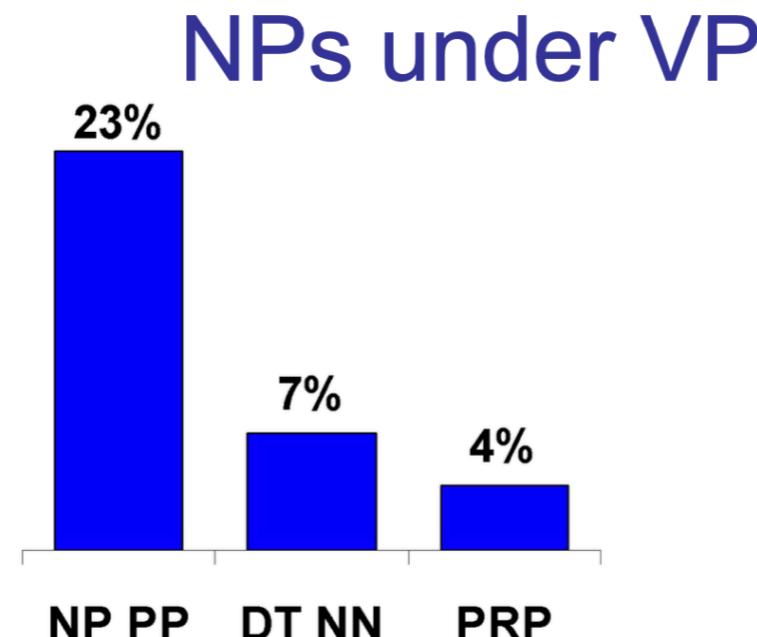
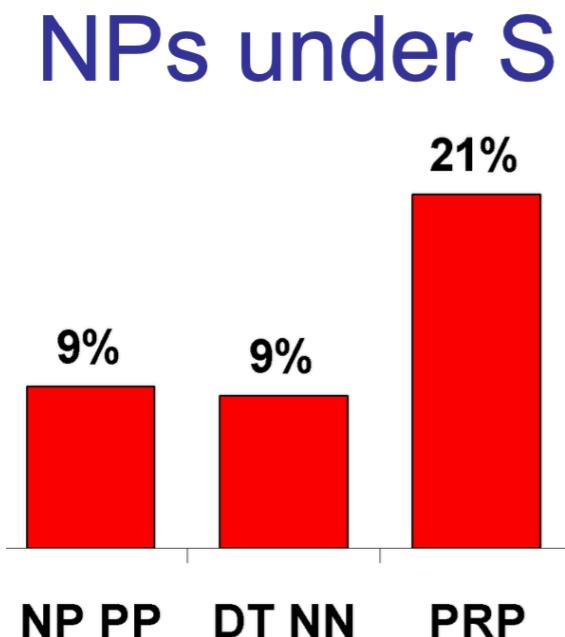


TABLE 1.

label	long name	example
NN	singular noun	pyramid
NNS	plural noun	lectures
NNP	proper noun	Khufu
VBD	past tense verb	claimed
VBZ	3rd person singular present tense verb	is
VBP	non-3rd person singular present tense verb	have
VBN	past participle	found
PRP	pronoun	they
PRP\$	possessive pronoun	their
JJ	adjective	public
IN	preposition complementizer	in that
DT	determiner	the

# Statistical NLP

- Like most other parts of AI, NLP is dominated by statistical methods
  - Typically more robust than earlier rule-based methods
  - Relevant statistics/probabilities are *learned from data*
  - Normally requires lots of data about any particular phenomenon
  -

# Statistical NLP

- Sparse data due to Zipf's Law
  - To illustrate, let's look at the frequencies of different words in a large text corpus
  - Assume "word" is a string of letters separated by spaces

# Statistical NLP

- Most frequent words in the English Europarl corpus (out of 24m word tokens)

<b>any word</b>		<b>nouns</b>	
Frequency	Token	Frequency	Token
1,698,599	the	124,598	European
849,256	of	104,325	Mr
793,731	to	92,195	Commission
640,257	and	66,781	President
508,560	in	62,867	Parliament
407,638	that	57,804	Union
400,467	is	53,683	report
394,778	a	53,547	Council
263,040	I	45,842	States

# Word Counts: Raw Words

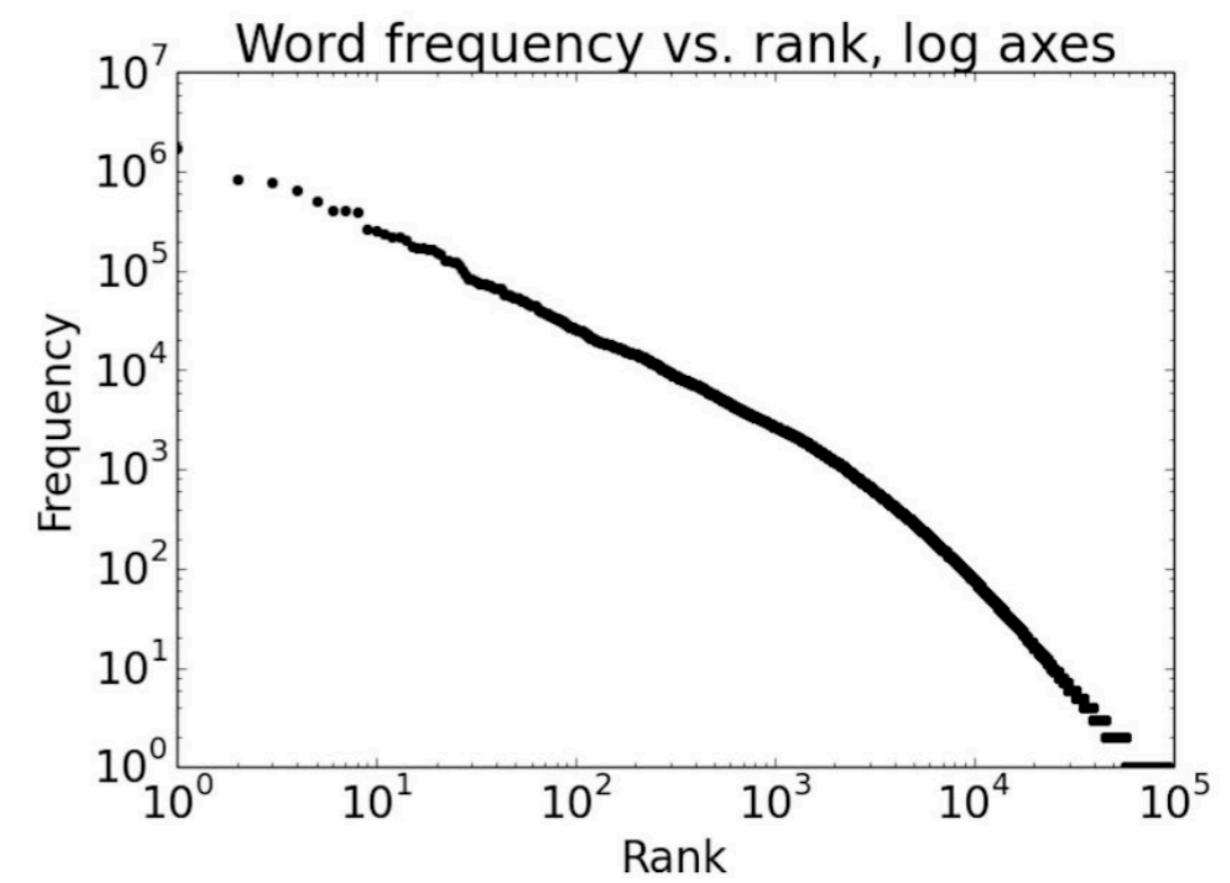
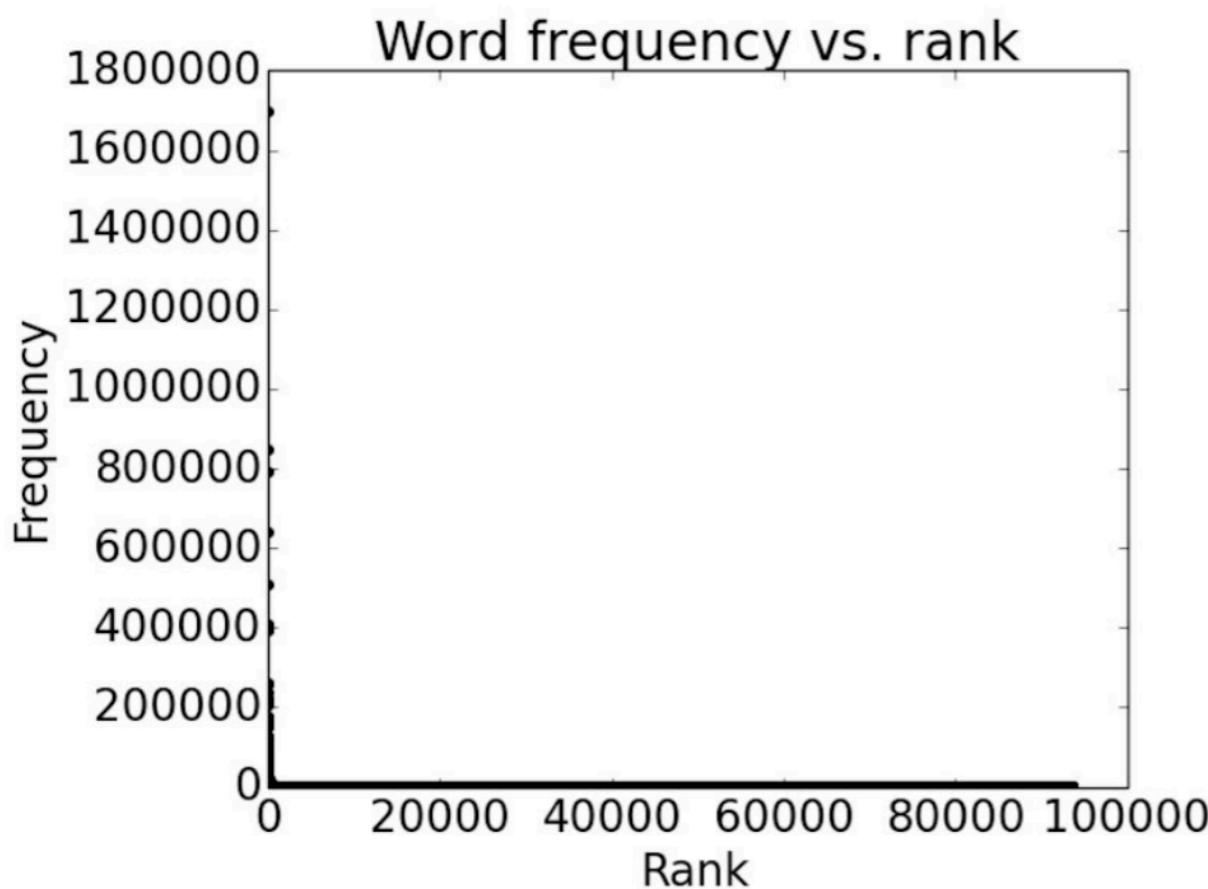
But also, out of 93,638 distinct words (word types), 36,231 occur only once.

Examples:

- cornflakes, mathematicians, fuzziness, jumbling
- pseudo-rapporteur, lobby-ridden, perfunctorily,
- Lycketoft, UNCITRAL, H-0695
- policyfor, Commissioneris, 145.95, 27a

# Plotting Word Frequencies

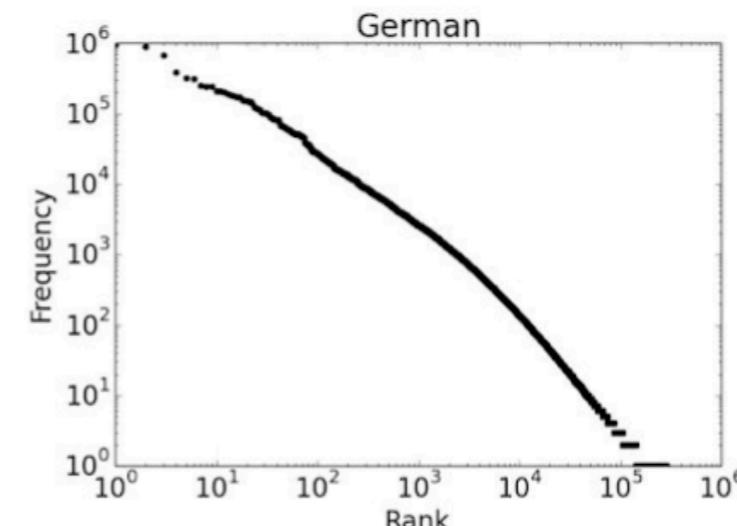
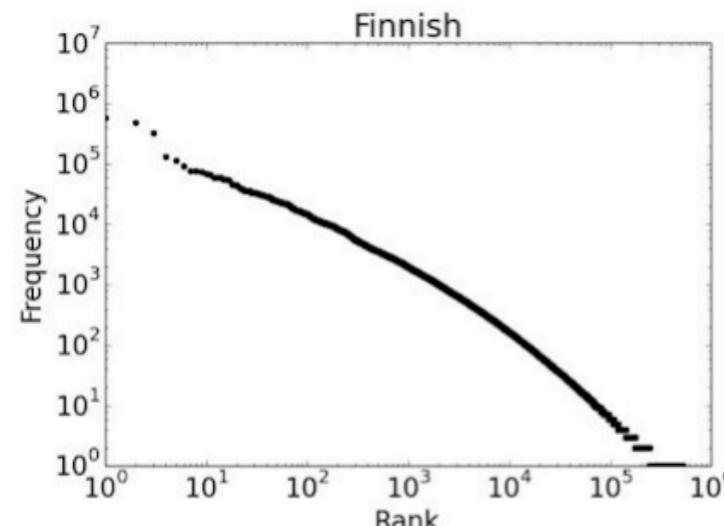
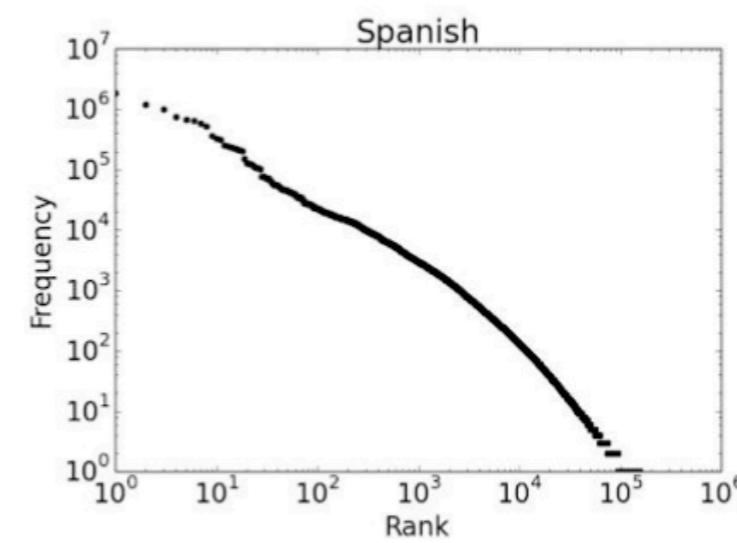
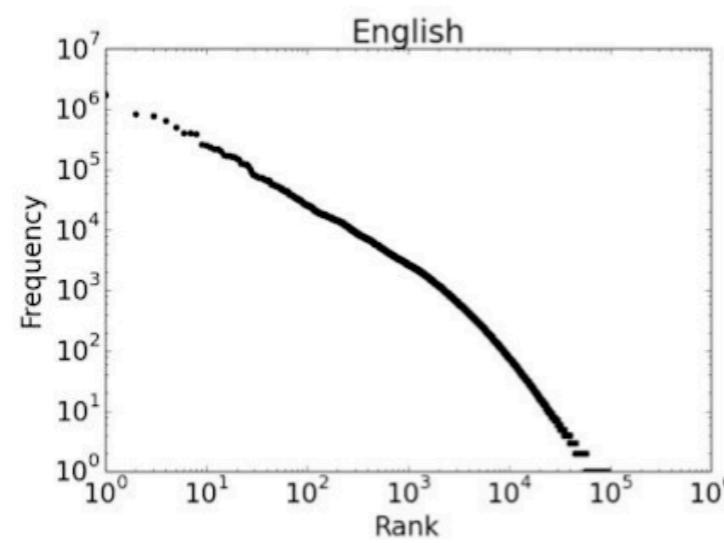
- Order words by frequency. What is the frequency of  $n_{th}$  ranked word?



# Zipf's Law

## Implications:

- Regardless of how large our corpus is, there will be a lot of infrequent (and zero-frequency!) words
- We need to find clever ways to estimate probabilities for things we have rarely or never seen

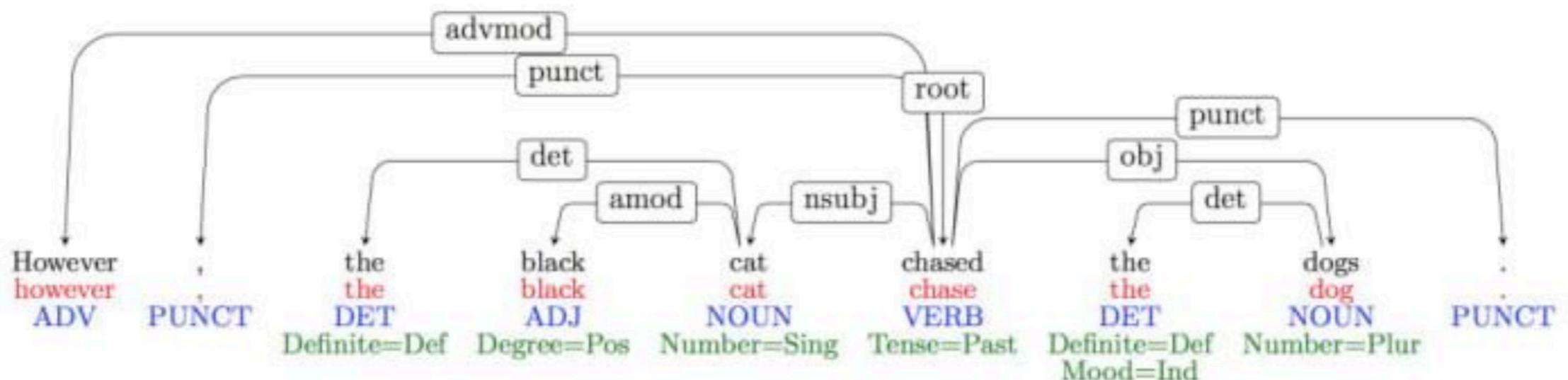


# Why NLP is Hard?

- Ambiguity
- Scale
- Sparsity
- Variation
- Expressivity
- Unmodeled variables
- Unknown representations  $R$

# Variation

- Suppose we train a part of speech tagger or a parser on the Wall Street Journal

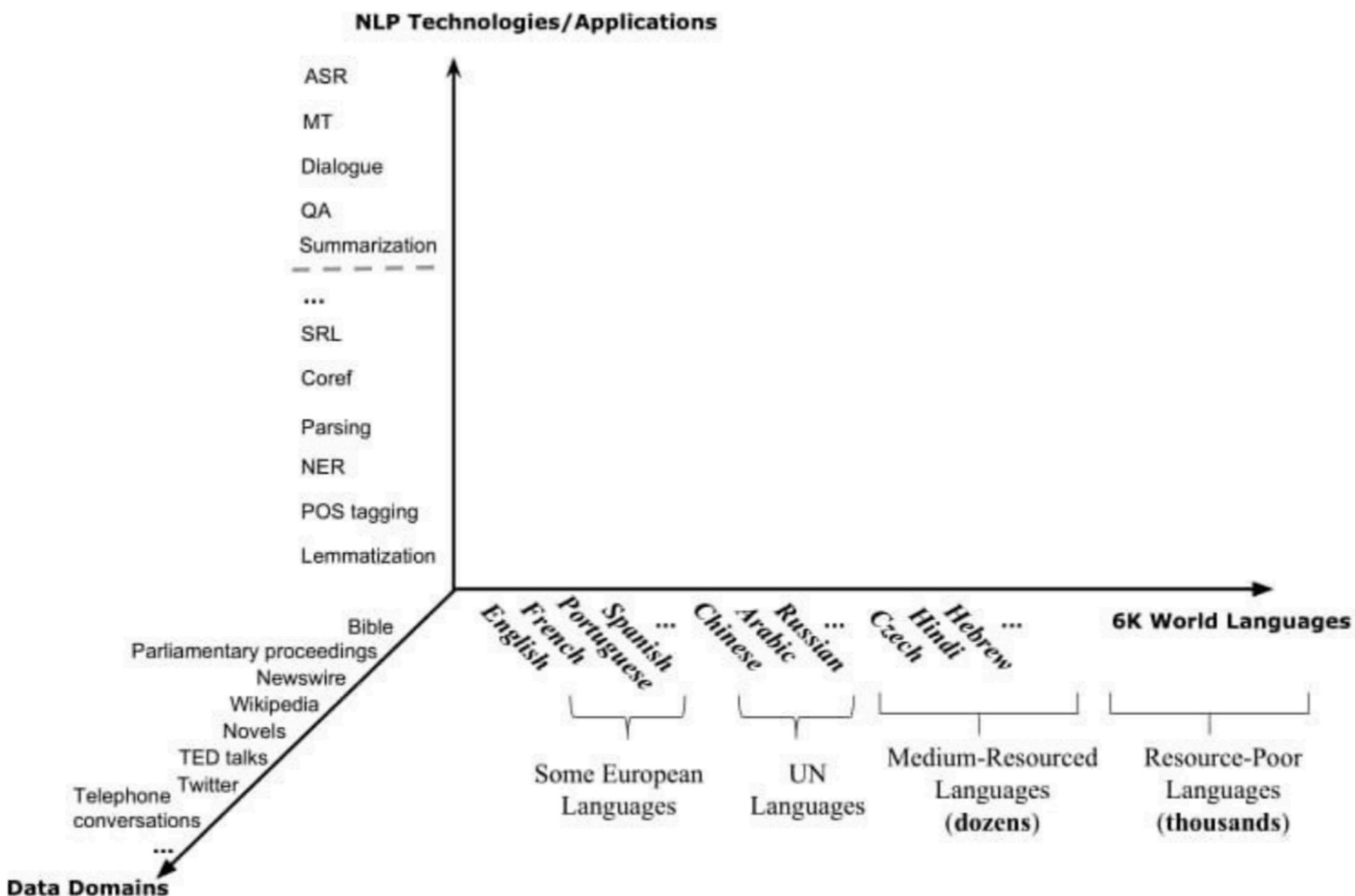


- What will happen if we try to use this tagger/parser for social media?

@\_rkpntrnte hindi ko alam babe eh, absent ako  
kanina I'm sick rn hahaha 😊🙌

# Variation

- Training data comes from diverse domains
- Potential distributional shift between train/test data



# Why NLP is Hard?

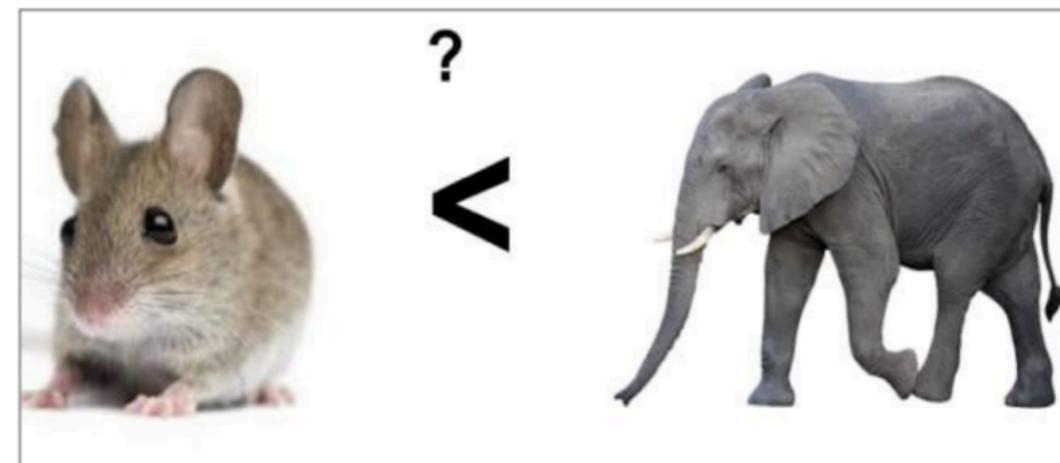
- Ambiguity
- Scale
- Sparsity
- Variation
- Expressivity
- Unmodeled variables
- Unknown representations  $\mathcal{R}$

# Expressivity

- Not only can one form have different meanings (ambiguity) but the same meaning can be expressed with different forms:
  - She gave the book to Tom vs. She gave Tom the book
  - Some kids popped by vs. A few children visited
  - Is that window still open? vs. Please close the window
  -

# Unmodeled Variables

- World knowledge
  - I dropped the glass on the floor and it broke
  - I dropped the hammer on the glass and it broke



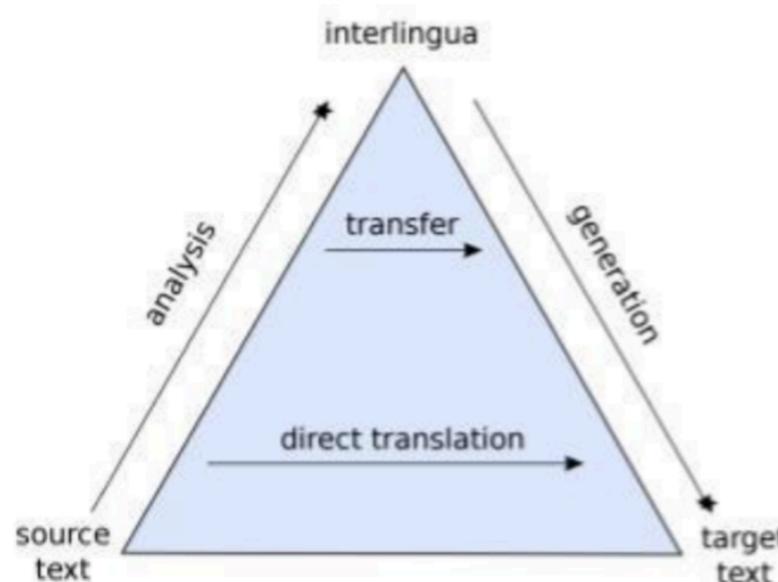
“Drink this milk”

# Unmodeled Meaning Representation

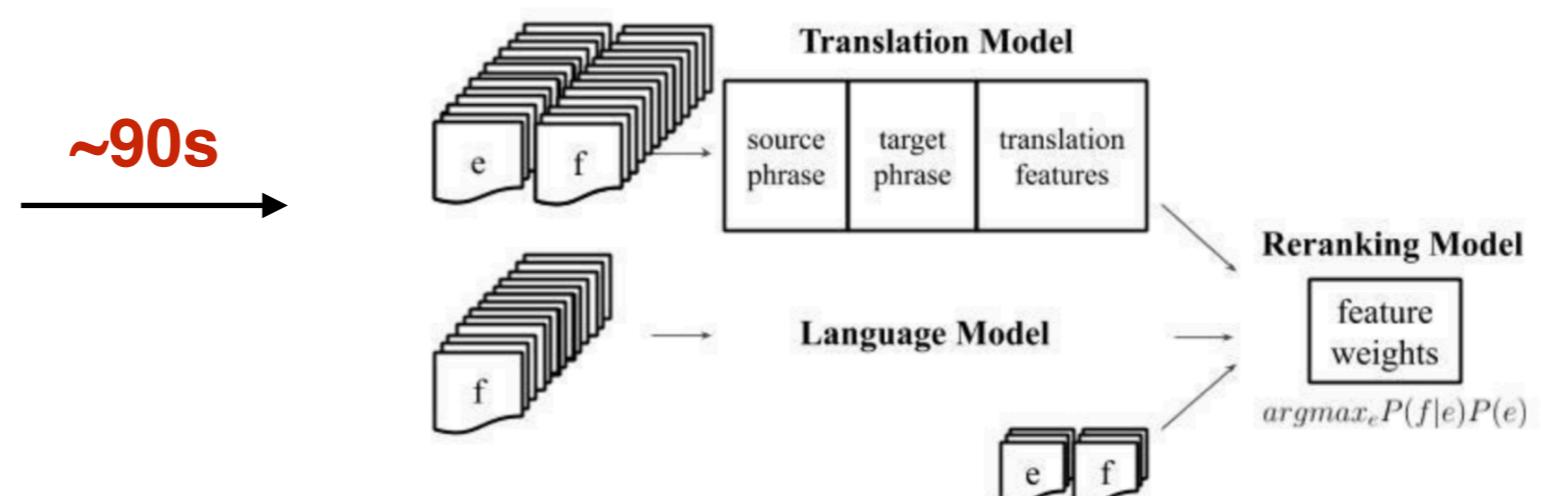
- Very difficult to capture what is  $\mathcal{R}$ , since we don't even know how to represent the knowledge a human has/needs:
  - What is the “meaning” of a word or sentence?
  - How to model context?
  - Other general knowledge?

# Symbolic and Probabilistic NLP

## Logic-based/Rule-based NLP

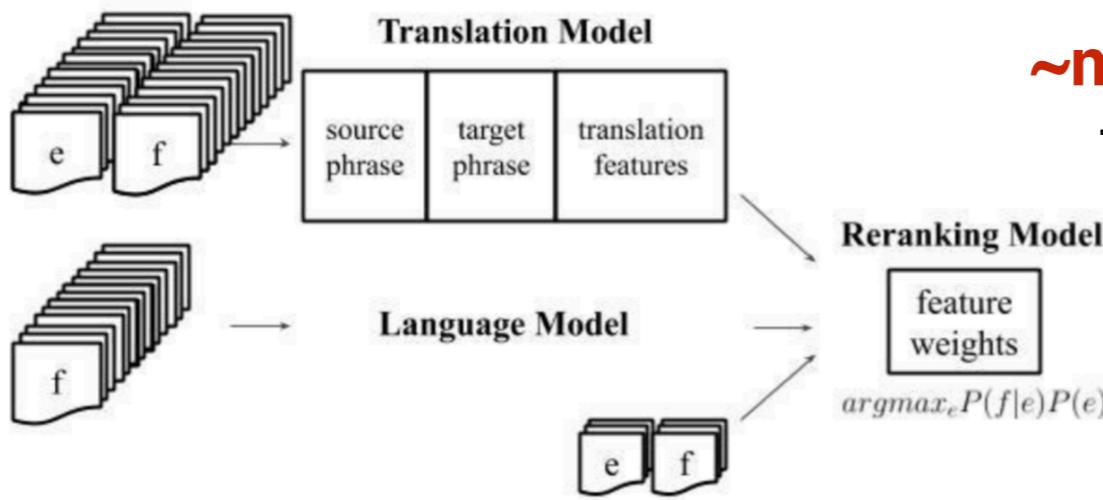


## Statistical NLP



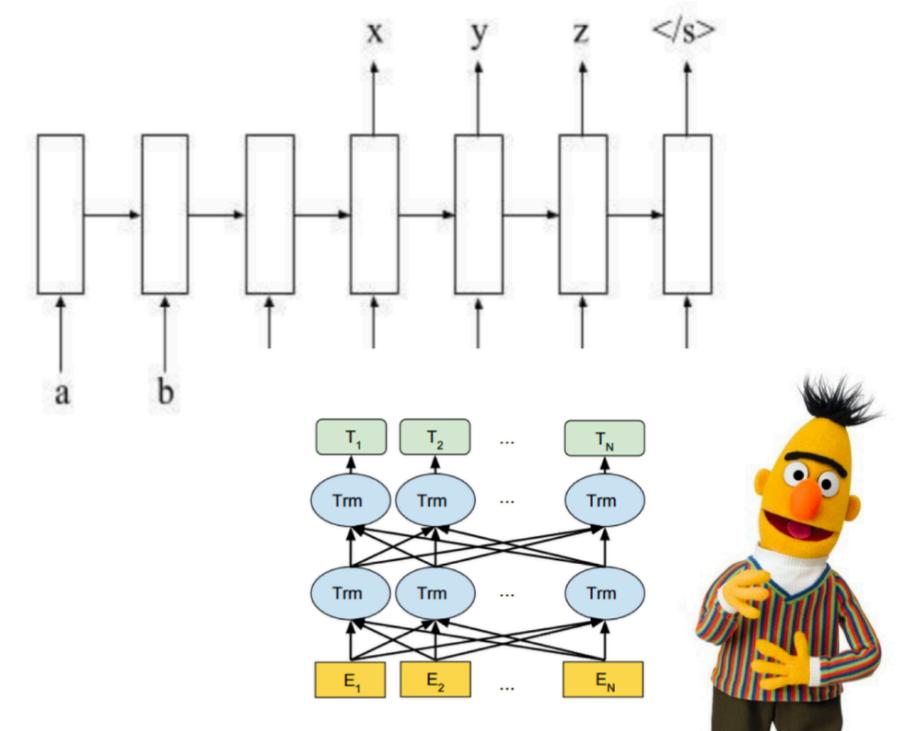
# Probabilistic and Connectionist NLP

## Engineered Features/Representations



~mid 2010s

## Learned Features/Representations



# NLP vs Machine Learning

- To be successful, a machine learner needs bias/assumptions; for NLP, that might be linguistic theory/representations.
- $\mathcal{R}$  is not directly observable.
- Symbolic, probabilistic, and connectionist ML have all seen NLP as a source of inspiring applications.

# NLP vs Linguistics

- NLP must process NL data as found in the world
- NLP  $\approx$  computational linguistics
- Linguistics has begun to use tools originating in NLP!

# Fields with Connections to NLP

- Machine learning
- Deep Learning
- Linguistics
- Cognitive science
- Information theory
- Data science
- Political science
- Psychology
- Economics
- Education

# NLP System Building Overview

# A General Framework for NLP Systems

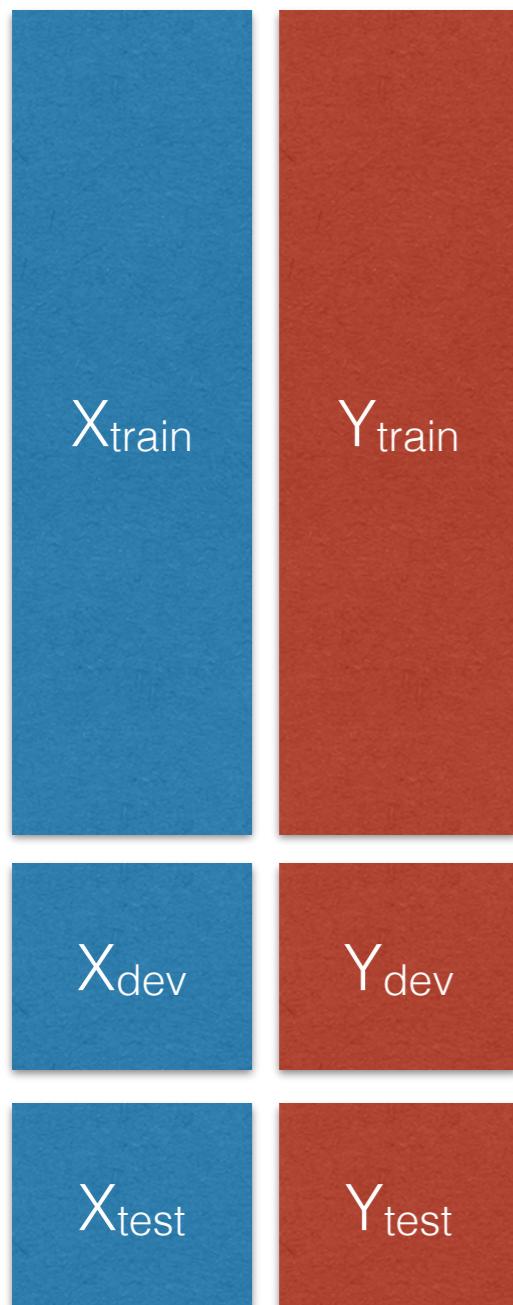
- Formally, create a function to map an **input  $X$  (*language*)** into an **output  $Y$** . Examples:

<u>Input <math>X</math></u>	<u>Output <math>Y</math></u>	<u>Task</u>
Text	Text in Other Language	Translation
Text	Response	Dialog
Text	Label	Text Classification
Text	Linguistic Structure	Language Analysis

- To create such a system, we can use
  - Manual creation of rules
  - Machine learning from paired data  $\langle X, Y \rangle$

# Train, Development, Test

- When creating a system, use three sets of data

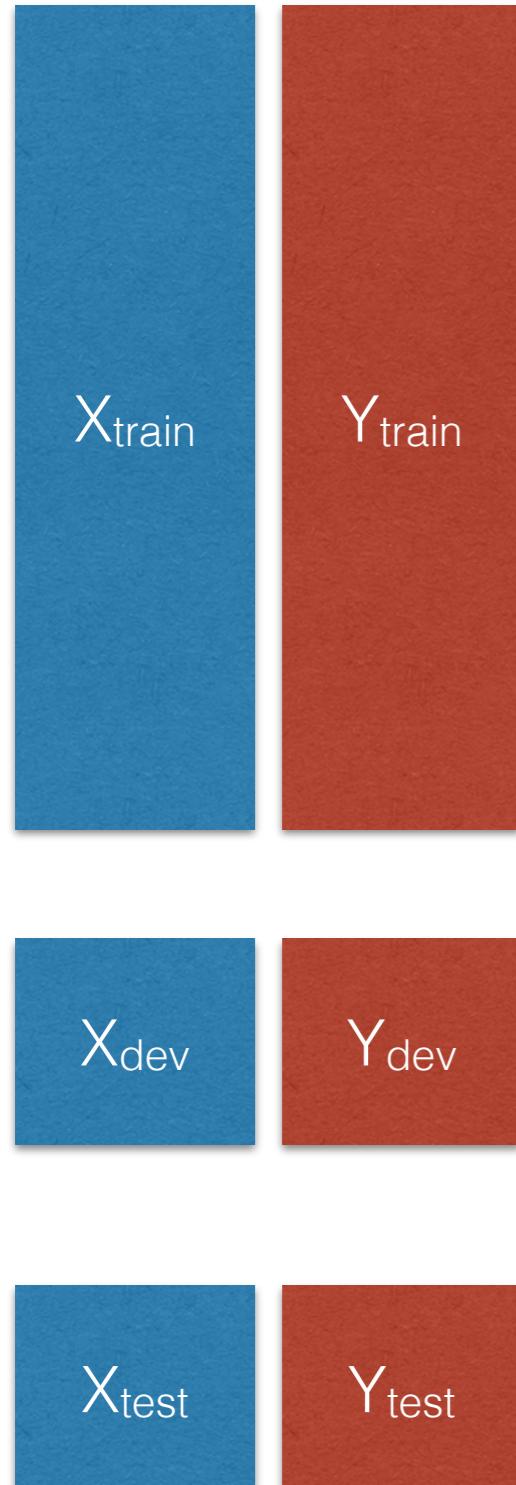


**Training Set:** Generally larger dataset, used during system design, creation, and learning of parameters.

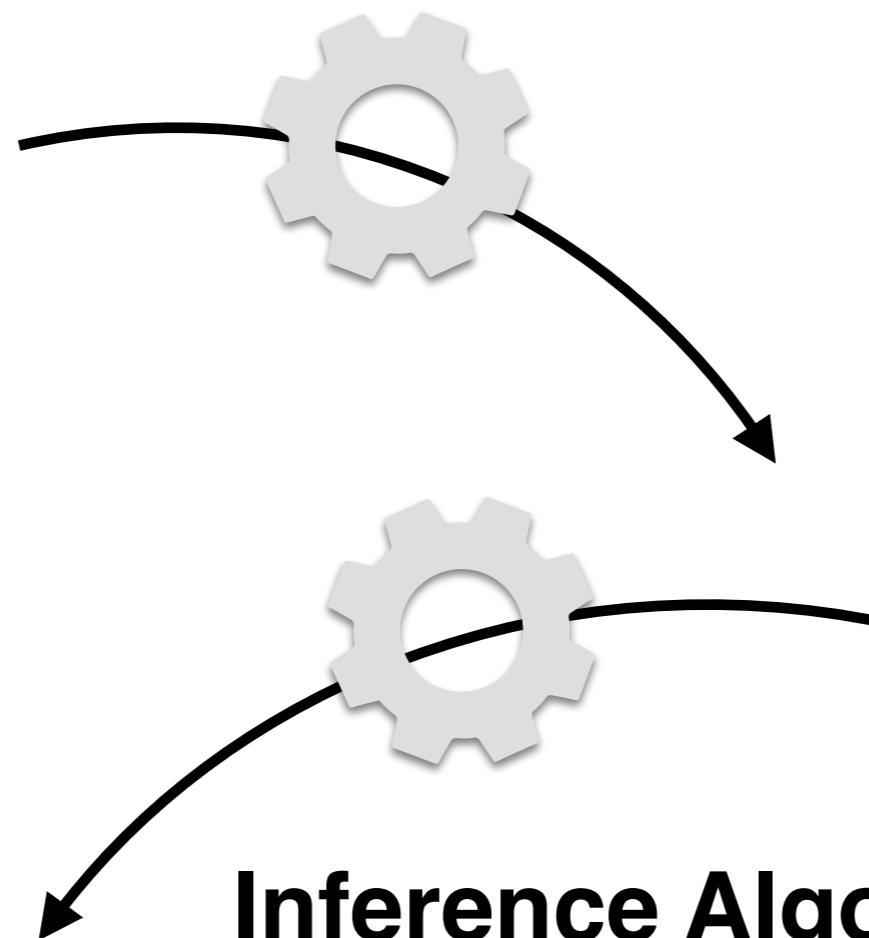
**Development ("dev", "validation") Set:** Smaller dataset for testing different design decisions ("hyper-parameters").

**Test Set:** Dataset reflecting the final test scenario, do not use for making design decisions.

# Machine Learning



## Learning Algorithm



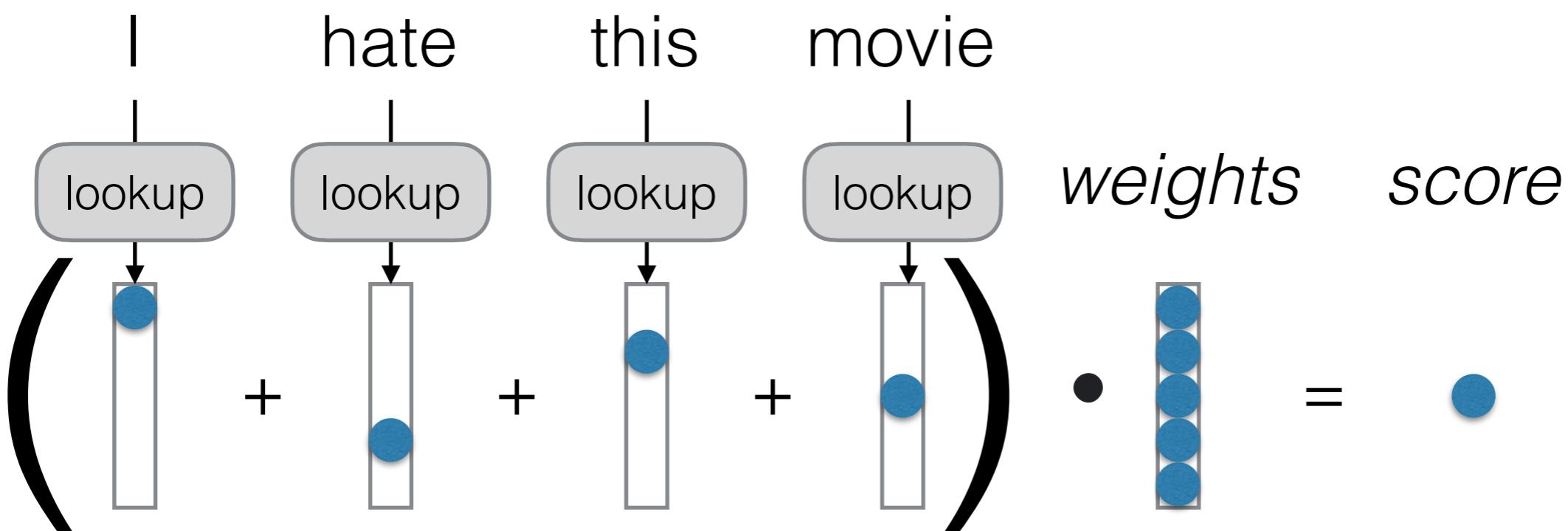
Learned  
Feature Extractor  $f$   
Scoring Function  $w$

$$\mathbf{h} = f(\mathbf{x})$$
$$s = \mathbf{w} \cdot \mathbf{h}$$

## Inference Algorithm

# Bag of Words (BOW)

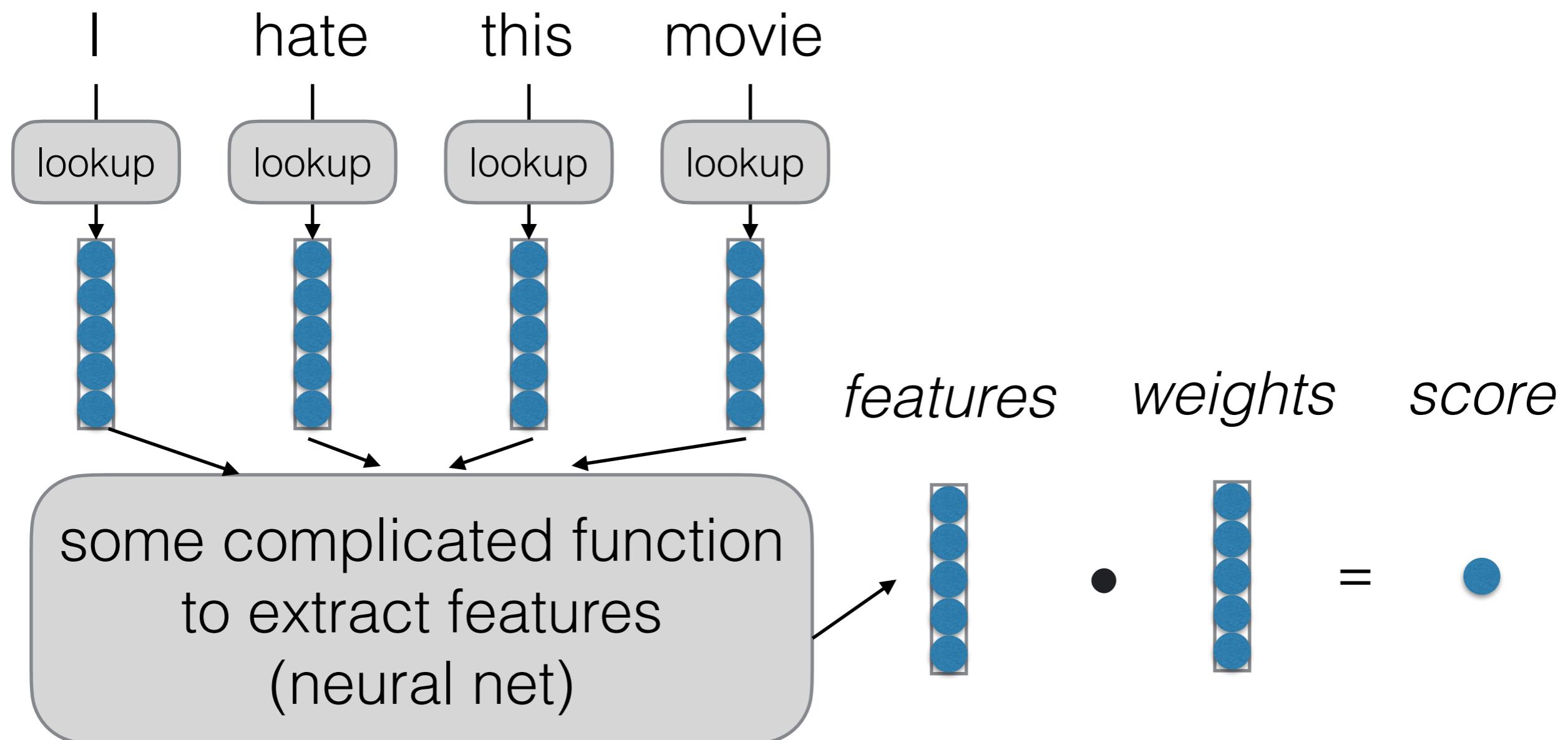
Convert each word into a one-hot vector:



Features  $f$  are based on word identity, weights  $w$  learned

Which problems mentioned before would this solve?

# Neural Network Models



# Class Goals

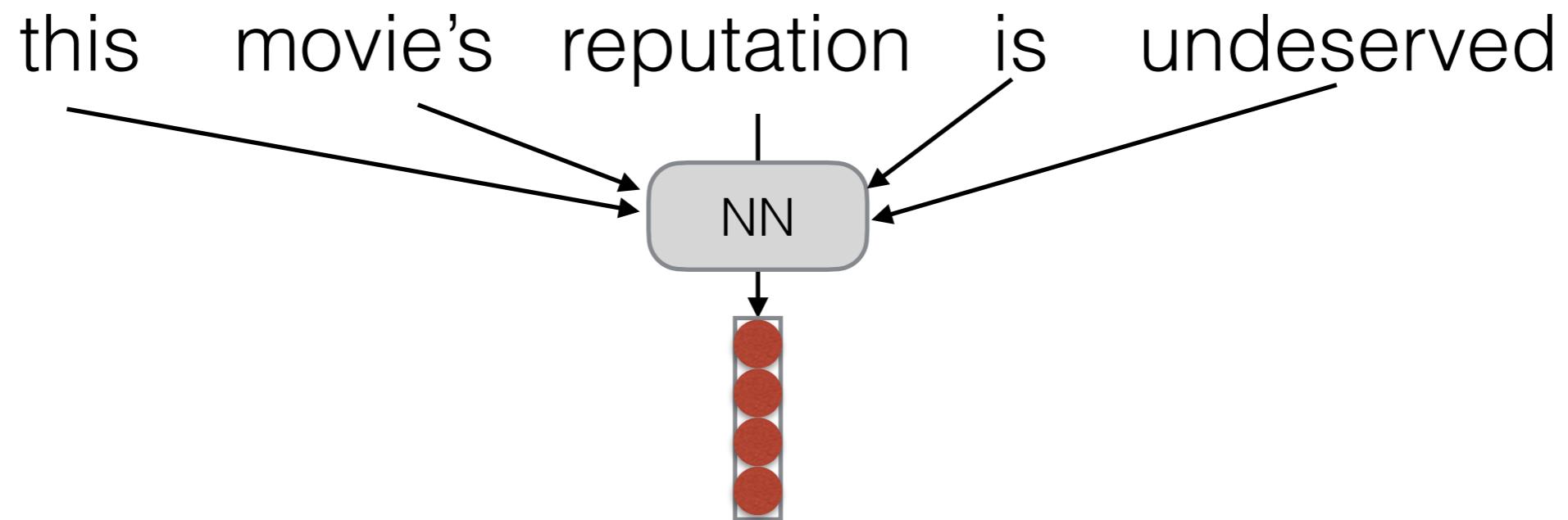
- Learn in detail about **building NLP systems from a research perspective**
- Learn basic and advanced topics in **machine learning and neural network approaches** to NLP
- Learn **basic linguistic knowledge** useful in NLP, and learn methods to **analyze linguistic structure**
- See several case studies of **NLP applications** and learn how to identify unique problems for each
- Learn how to debug **when and where NLP systems fail**, and build improvements based on this

# Prior Background

- There is no hard prerequisite to this course.
- But this is a **research-oriented** course. Here are some recommendations:
  - Take at least one **intro-level AI** course
  - Basic statistics/probability/linear algebra
  - Python programming, Deep Learning Library (e.g., PyTorch)

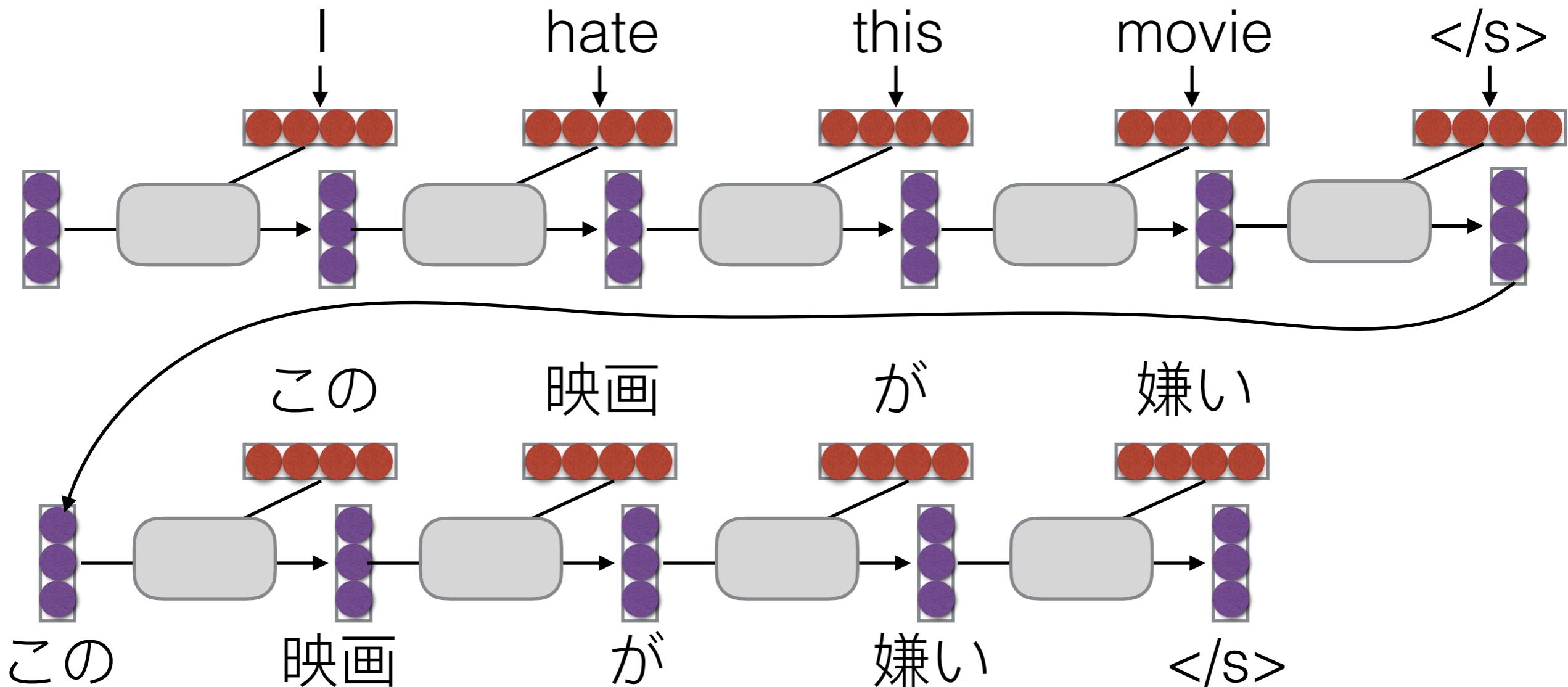
# Roadmap Going Forward

# Topic 1: NLP Fundamentals



- Text Classification and ML Fundamentals
- Language Modeling and NN Training Tricks
- Word Vectors
- Neural Network Basics and Toolkit Construction

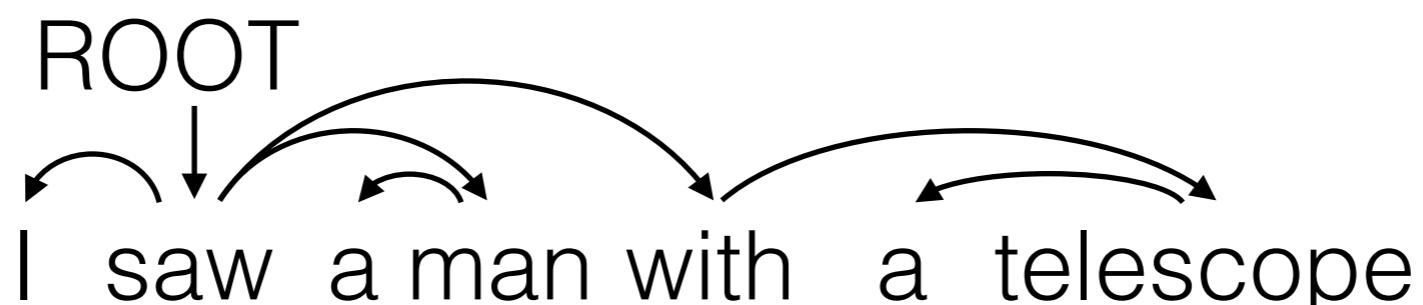
# Topic 2: Modeling and Neural Net Basics



- Recurrent Networks
- Conditioned Generation
- Attention

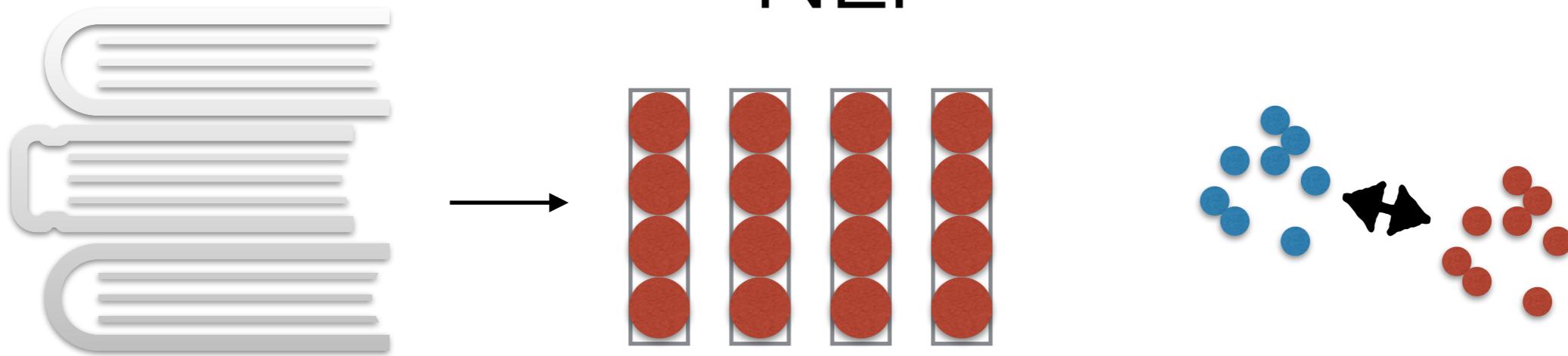
# Topic 3:

# Natural Language Analysis



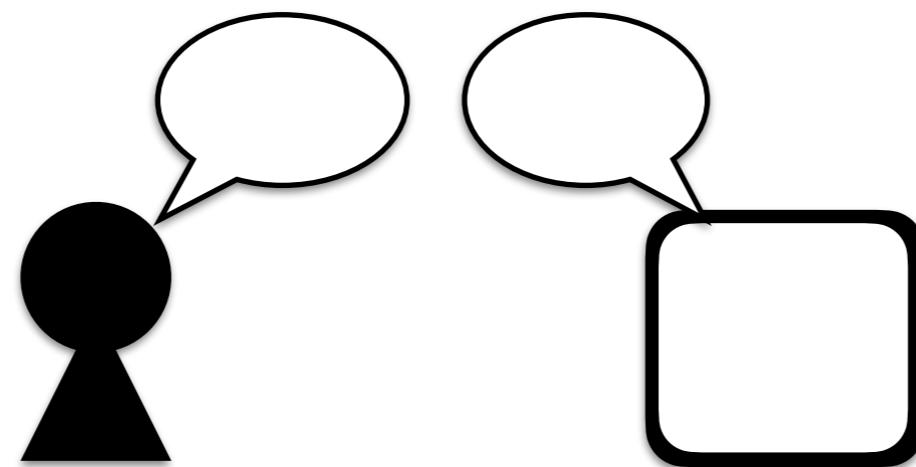
- Word Segmentation and Morphology
- Syntactic / Semantic Parsing

# Topic 4: Advanced Deep Learning Algorithms in NLP



- Pre-training and Self-supervised Learning
- Multi-task and Multi-lingual Learning
- Prompting and Few-shot Learning
- Long Sequence Models
- Structured Learning Algorithms
- Latent Variable Models

# Topic 5: NLP Applications



- Machine Reading QA
- Dialog
- Computational Social Science, Bias and Fairness
- Information Extraction and Knowledge-based QA
- NLP and multimodal machine learning

# Class Logistic

# Course Website

- <https://junjiehu.github.io/cs769-fall24/>

A screenshot of a web browser window displaying the course website. The address bar shows the URL: junjiehu.github.io/cs769-fall24/lectures/. The page title is "CS 769 NLP". On the right side of the header, there are links for "logistics", "lectures", and "home". Below the header, the word "Schedule" is prominently displayed in large, bold, dark font. Underneath it, a subtext says "All videos are available in this shared folder (Link)". A table below lists the course schedule with columns for Date, Lecture, Readings, and Logistics.

Date	Lecture	Readings	Logistics
<b>Module 1: Introduction to Natural Language Processing</b>			
9/4	Lecture #1 : <b>Overview of NLP</b> [ <a href="#">slides</a>   <a href="#">video</a> ]	<ul style="list-style-type: none"><li>• SLP3 ch2</li></ul>	
9/9	Lecture #2 : <b>Text Classification</b> [ <a href="#">slides</a>   <a href="#">video</a> ]	<ul style="list-style-type: none"><li>• SLP3 ch4</li><li>• E ch2-4</li><li>• Dani et al. 2017</li></ul>	
9/11	Lecture #3 : <b>Deep Learning Basics</b> [ <a href="#">slides</a>   <a href="#">video</a> ]	<ul style="list-style-type: none"><li>• Colab example</li><li>• Gradient Descent (Ruder 2016)</li><li>• Dropout (Srivastava et al. 2014)</li></ul>	

## Schedule

All videos are available in this shared folder ([Link](#)).

Date	Lecture	Readings	Logistics
<b>Module 1: Introduction to Natural Language Processing</b>			
9/4	Lecture #1 : <b>Overview of NLP</b> [ <a href="#">slides</a>   <a href="#">video</a> ]	<ul style="list-style-type: none"><li>• SLP3 ch2</li></ul>	
9/9	Lecture #2 : <b>Text Classification</b> [ <a href="#">slides</a>   <a href="#">video</a> ]	<ul style="list-style-type: none"><li>• SLP3 ch4</li><li>• E ch2-4</li><li>• Dani et al. 2017</li></ul>	
9/11	Lecture #3 : <b>Deep Learning Basics</b> [ <a href="#">slides</a>   <a href="#">video</a> ]	<ul style="list-style-type: none"><li>• Colab example</li><li>• Gradient Descent (Ruder 2016)</li><li>• Dropout (Srivastava et al. 2014)</li></ul>	

# Class Delivery Format: In Person

- Office hour section remains **online on Zoom**  
**(<https://uwmadison.zoom.us/j/95113800299>)**
- Contact the instructor if there's any concerns

# Class Content Format

- **Before class:** For some classes, do recommended reading
- **During class:**
  - *Lecture/Discussion:* Go through material and discuss
  - *Code/Data Walk:* The instructor will sometimes walk through some demonstration code, data, or model predictions

# Assignments

- **Individual** assignments (Assignments 1, 2, 4)
  - Each individual will have **three late days** in total for the three individual assignments.
- **Group** assignments (Assignments 3, 5)
  - Each group should consist of **three members**. Group of two are only permitted when the class size cannot be divided by 3.
  - Another **three late days** are given to the group.

# Assignments

- **Assignment 1 - Text Classifier / Questionnaire:** *Individually* implement a text classifier, and indicate your project topic (15%)
- **Assignment 2 - Text Classifier with Pre-trained LM:** *Individually* implement a text classifier (15%)
- **Assignment 3 - Project Proposal (SOTA Survey / Re-implementation):** Re-implement and reproduce results from a recently published NLP paper, and proposal new ideas (20%)
- **Assignment 4 - Beam Search:** Individually implement beam search for text generation. (20%)
- **Assignment 5 - Final Project:** Perform a unique project that either (1) improves on state-of-the-art, or (2) applies NLP models to a unique task. Have an oral presentation and write a report. (30%)

# Instructors

- **Instructor:**
  - Junjie Hu
- **Grader:**
  - TBD
- **Website:** <https://junjiehu.github.io/cs769-fall24/>
- **Piazza:** <https://piazza.com/wisc/fall2024/cs769>
- **Canvas:** <https://canvas.wisc.edu/courses/426756/>
- **Assignments:** <https://github.com/JunjieHu/cs769-assignments/>

Thanks, Any Questions?