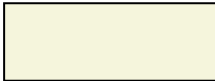# Natural Language Processing (NLP)

Data and Knowledge

- Classic knowledge representation worry: How will a machine ever know that…

- Ice is frozen water?
- Beige looks like this:
- Chairs are solid?

- Answers:

- 1980: write it all down
- 2000: get by without it
- 2020: learn it from data

Brief history of NLP

- Foundational Insights: 1940s and 1950s
  - Two foundational paradigms
    - Automaton
    - Probabilistic/Information-Theoretic models

- The two camps: 1957-1970
  - Symbolic paradigm: the work of Chomsky and others on formal language theory and generative syntax (1950s ~ mid 1960s)
  - Stochastic paradigm
    - In departments of statistics

Brief history of NLP

- Four paradigms: 1970-1983, explosion in research in speech and language processing
    - Stochastic paradigm
    - Logic-based paradigm
    - Natural language understanding
    - Discourse modeling paradigm
- Empiricism and Finite State Models Redux: 1983-1993

Brief history of NLP

- The Fields Comes Together: 1994-1999
    - Probabilistic and data-driven models had become quite standard
- The Rise of Machine Learning: 2000-now
    - Large amount of spoken and textual data become available
    - Widespread availability of high-performance computing systems

NLP Introduction

✖ **What is NLP?**

✖ NLP problem can be divided into two tasks:

+ **Processing written text**, using **lexical, syntactic and semantic knowledge** of the language as well as the required real world information.

+ **Processing spoken language**, using all the information needed above **plus** additional knowledge about **phonology** as well as enough added information to handle the further ambiguities that arise in speech.

NLP Introduction

✖ The Problem : English sentences are incomplete descriptions of the information that they are intended to convey.

Example: **"Some dogs are outside."** is incomplete – it can mean

✚ Some dogs are on the lawn.

✚ Three dogs are on the lawn.

✚ Moti, Hira & Buzo are on the lawn.

NLP Introduction

✖ The Problem : The same expression means different things in different context.

➕ Where's the water? (In a lab, it must be pure)

➕ Where's the water? (When you are thirsty, it must be potable or drinkable)

➕ Where's the water? (Dealing with a leaky roof, it can be filthy)

NLP Introduction

✖ The problem : There are lots of ways to say the same thing :

✚ Mary was born on October 11.
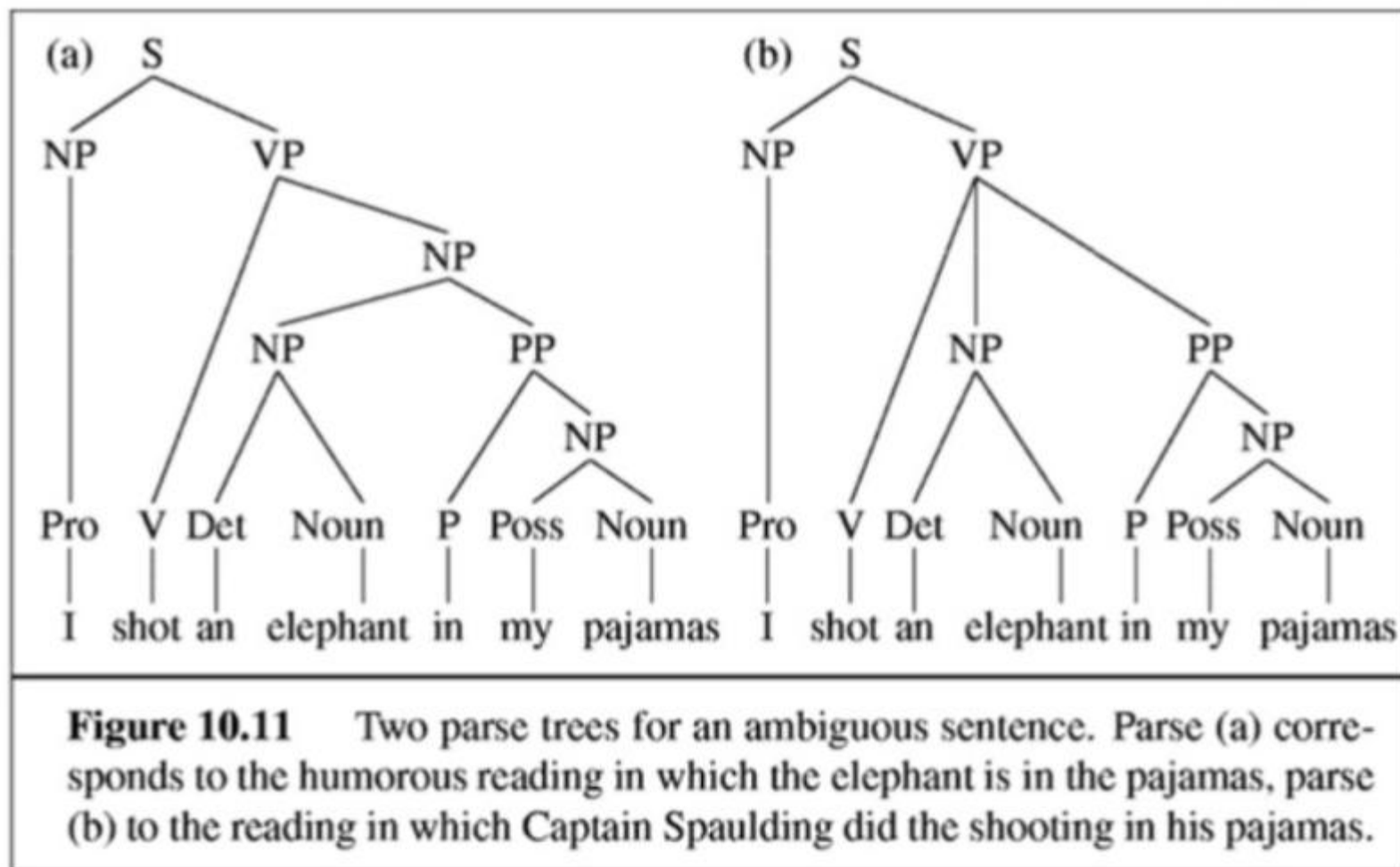
✚ Mary's birthday is October 11.

Steps in NLP

1. Morphological Analysis:

- Individual words are analyzed into their components and non-word tokens such as punctuation are separated from the words.

- Tries to extract **root** word from **decline or inflectional** form of word after removing suffices and prefixes. Ex: getting the root "push " from declined from pushes, pushed, pushing, etc.

- Assign appropriate syntactic categories such as **noun, verb, adjective** etc. to all words in the sentence.

Steps in NLP

2. Syntactic Analysis:

- Use the result of Morphological analysis to build a structure description of sentence based on grammatical rules. This step is called *parsing.*

- Creating a parse tree is the first step towards understanding a sentence.

Figure 10.11 Two parse trees for an ambiguous sentence. Parse (a) corresponds to the humorous reading in which the elephant is in the pajamas, parse (b) to the reading in which Captain Spaulding did the shooting in his pajamas.

3. **Semantic Analysis**:

- The structures created by the syntactic analyzer are assigned **meanings.**

- It **maps** individual words in to corresponding object in the knowledge base and combine the words with each other with semantic rules.

4. Discourse integration:

- The meaning of an individual sentence may depend on the sentences that **precede** it and may influence the meanings of the sentences that **follow** it.

- Ex: the word **"it"** in sentence "John wanted **it"**, depends up on **prior** discourse context.

5. Pragmatic Analysis:

- It refers to **intended** meaning of sentences used in different contexts. The context affects the interpretation of the sentence.

  Ex: (a) Don't put the cart before the horse.

  (b) I saw a boy with binoculars in the garden.