

# NLP

# Introduction to NLP

*Why is NLP hard?*

# Example

*Time flies like an arrow.*

- How many different interpretations does the above sentence have?
- How many of them are reasonable/grammatical?

## Quiz Answer

- The most obvious meaning is
  - time flies very fast; as fast as an arrow.
- This is a metaphorical interpretation.
  - Computers are not really good at metaphors.
- Other interpretations:
  - Flies like honey → flies like an arrow → fruit flies like an arrow
  - Take a stopwatch and time the race → time the flies

# More Classic Examples

- Beverly Hills
- Beverly Sills
- The box is in the pen
- The pen is in the box
- Mary and Sue are mothers
- Mary and Sue are sisters
- Every American has a mother
- Every American has a president
- We gave the monkeys the bananas because they were hungry
- We gave the monkeys the bananas because they were over-ripe
  - <http://specgram.com/CLIII.4/08.phlogiston.cartoon.zhe.html>

# Syntax vs. Semantics

*\* Little a has Mary lamb.*

*? Colorless green ideas sleep furiously.*

[Chomsky 1957]

## Ambiguous Words

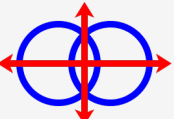
- ball, board, plant
  - meaning
- fly, rent, tape
  - part of speech
- address, resent, entrance, number, unionized
  - pronunciation – give it a try

## Answer To The Quiz

- **address**
  - The stress can be on either syllable. Compare with transport, effect, outline
- **resent**
  - As a verb infinitive or as “re-sent” a letter
- **entrance**
  - As a noun or as a verb meaning to put someone in a trance
- **number**
  - As a noun but also as the comparative of the adjective “numb”



# Ambiguity

- Not in computer languages (by design)!
- Or Lojban 
- Noun–noun phrases: (XY)Z vs. X(YZ)
  - *science fiction writer*
  - *customer service representative*
  - *state chess tournament*

## NACLO Problems

- One Two Tree, by Noah Smith, Kevin Gimbel, and Jason Eisner
  - <http://www.naclo.cs.cmu.edu/problems2012/N2012-R.pdf>
- Fakepapershellmaker, by Willie Costello
  - <http://www.naclo.cs.cmu.edu/problems2008/N2008-F.pdf>

# NACLO Problem Solutions

- One Two Tree
  - <http://www.naclo.cs.cmu.edu/problems2012/N2012-RS.pdf>
- Fakepapershelfmaker
  - <http://www.naclo.cs.cmu.edu/problems2008/N2008-FS.pdf>

# Types of Ambiguity 1/2

- **Morphological:**
  - Joe is quite impossible. Joe is quite important.
- **Phonetic:**
  - Joe's finger got number.
- **Part of speech:**
  - Joe won the first round.
- **Syntactic:**
  - Call Joe a taxi.
- **Pp attachment:**
  - Joe ate pizza with a fork / with meatballs / with Samantha / with pleasure.
- **Sense:**
  - Joe took the bar exam.
- **Modality:**
  - Joe may win the lottery.

# Types of Ambiguity 2/2

- **Subjectivity:**
  - Joe believes that stocks will rise.
- **Cc attachment:**
  - Joe likes ripe apples and pears.
- **Negation:**
  - Joe likes his pizza with no cheese and tomatoes.
- **Referential:**
  - Joe yelled at Mike. He had broken the bike.
  - Joe yelled at Mike. He was angry at him.
- **Reflexive:**
  - John bought him a present.
  - John bought himself a present.
- **Ellipsis and parallelism:**
  - Joe gave Mike a beer and Jeremy a glass of wine.
- **Metonymy:**
  - Boston called and left a message for Joe.

# Other Sources of Difficulties

- Non-standard, slang, and novel words and usages
  - A360, 7342.67, +1-646-555-2223
  - “spam” or “friend” as verbs
  - yolo, selfie, chillax – recently recognized as dictionary words
  - [www.urbandictionary.com](http://www.urbandictionary.com) – (Parental Warning!)
- Inconsistencies
  - junior college, college junior
  - pet spray, pet llama
- Typoses and grammatical errors ☺
  - Receipt, John Hopkins, should of
- Parsing problems
  - Cup holder
  - Federal Reserve Board Chairman

## Other Sources of Difficulties

- Complex sentences
- Counterfactual sentences
- Humor and sarcasm
- Implicature/inference/world knowledge:
  - I was late because my car broke down.
  - Implies I have a car, I use the car to get to places, the car has wheels, etc.
  - What is not explicitly mentioned, what is world knowledge?
- Semantics vs. pragmatics
  - Do you know the time?
- Language is hard even for humans (both L1 and L2)

# Synonyms and Paraphrases

The S&P 500 climbed 6.93, or 0.56 percent, to 1,243.72, its best close since June 12, 2001.

The Nasdaq gained 12.22, or 0.56 percent, to 2,198.44 for its best showing since June 8, 2001.

The DJIA rose 68.46, or 0.64 percent, to 10,705.55, its highest level since March 15.



# Synonyms and Paraphrases

The S&P 500 climbed 6.93, or 0.56 percent, to 1,243.72, its best close since June 12, 2001.

The Nasdaq gained 12.22, or 0.56 percent, to 2,198.44 for its best showing since June 8, 2001.

The DJIA rose 68.46, or 0.64 percent, to 10,705.55, its highest level since March 15.

# NLP