

# I256: Applied Natural Language Processing

Marti Hearst  
Lecture 1

# Course Instructors

Marti Hearst



John Semerdjian



... and you!



# Today

- What Makes Language Special?
- NLP Applications and the State of the Art
- Intended Learning Outcomes
- Administrivia

# What is NLP?



- Fundamental goal: deep understand of *broad language*
  - Not just string processing or keyword matching!
  - But ... along the way we develop a lot of useful technology.

# Which Fields Border NLP?

## ■ (Computational) Linguistics

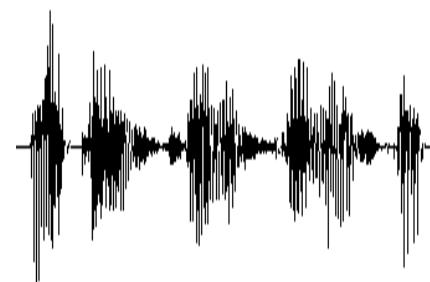
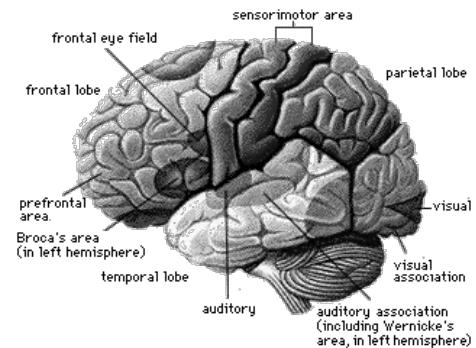
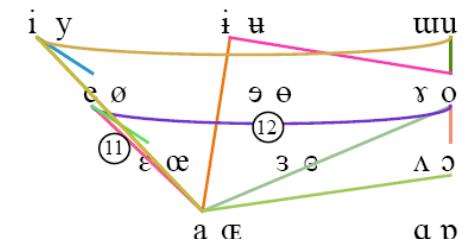
- Using computational methods to learn more about how language works
  - Use linguistics to understand the structure of language.

## ■ Cognitive Science

- Figuring out how the human brain works
  - Includes the bits that do language
  - Humans: the only working NLP prototype!

## ■ Speech Processing

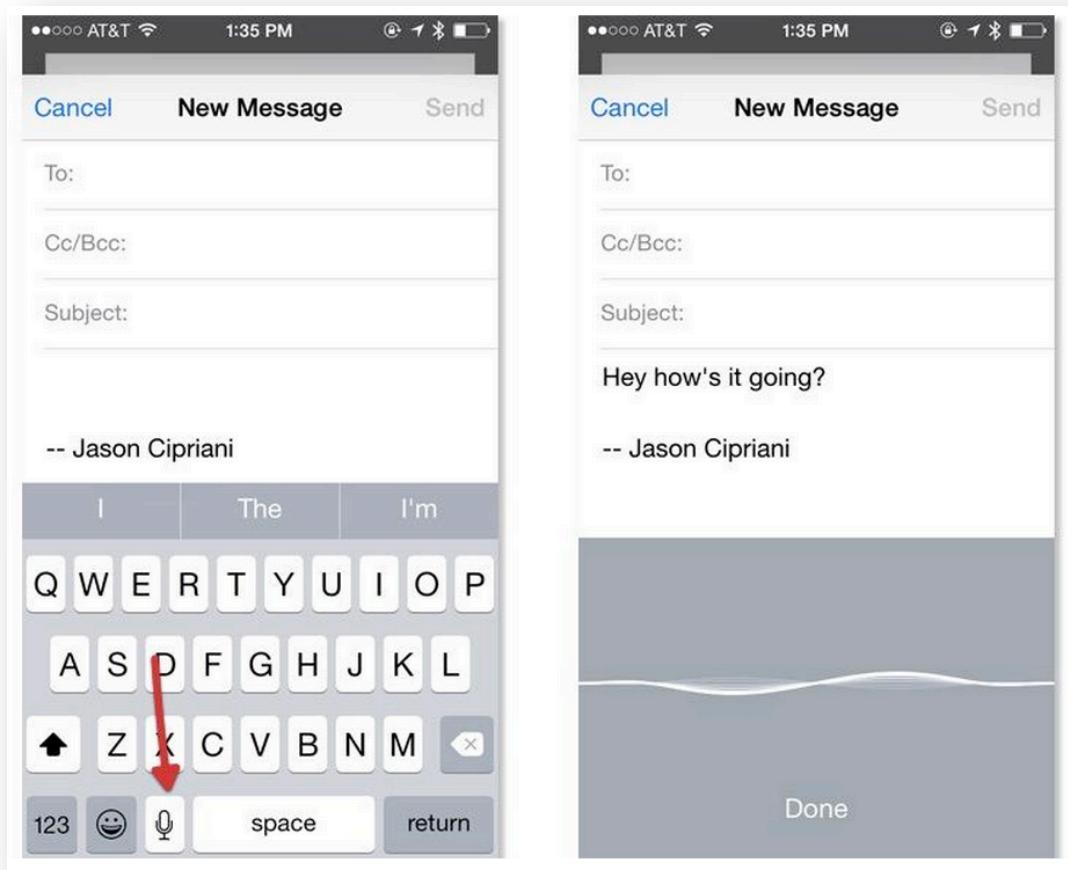
- Mapping audio signals to text
  - Two main components: acoustic models and language models.





**IF YOU'D SHOWN ME THESE IN  
GRAD SCHOOL, YOU WOULD  
HAVE BLOWN MY MIND...**

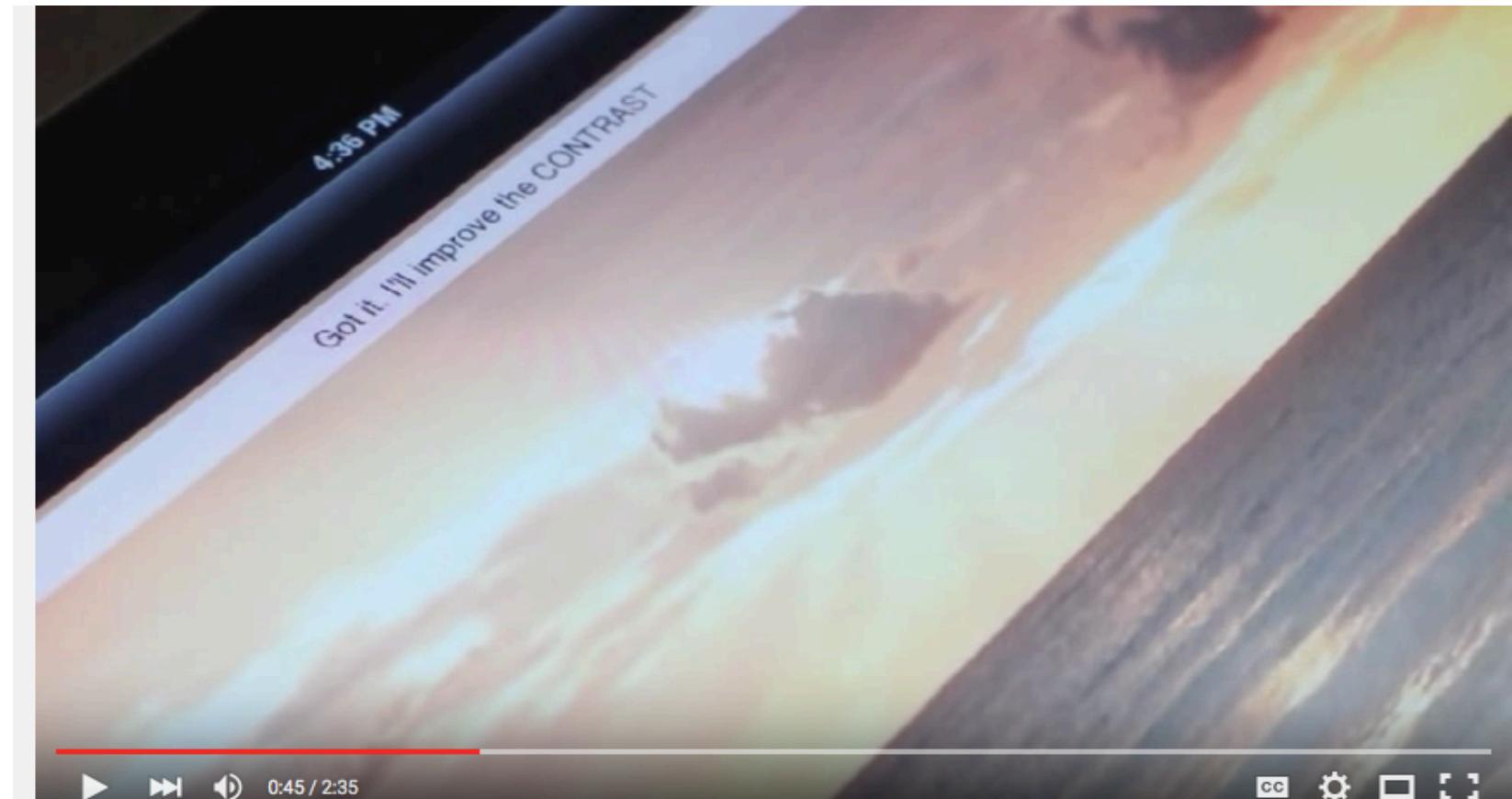
# Have My Phone Recognize My Speech



# Get Turn by Turn Voice Directions with Siri



# I Can Talk Intelligently with PhotoShop



# I Can Show the Computer a Photo, Have it Write Out a Sentence



Human: "Three different types of pizza on top of a stove."

Computer model: "A pizza sitting on top of a pan on top of a stove."

Vinyals, Toshev Bengio, Erhan, 2014  
<http://arxiv.org/pdf/1411.4555v2.pdf>

A person riding a motorcycle on a dirt road.



A group of young people playing a game of frisbee.



A herd of elephants walking across a dry grass field.



Two dogs play in the grass.



Two hockey players are fighting over the puck.



A close up of a cat laying on a couch.



A skateboarder does a trick on a ramp.



A little girl in a pink hat is blowing bubbles.



A red motorcycle parked on the side of the road.



A dog is jumping to catch a frisbee.



A refrigerator filled with lots of food and drinks.



A yellow school bus parked in a parking lot.



Describes without errors

Describes with minor errors

Somewhat related to the image

Unrelated to the image

Figure 5. A selection of evaluation results, grouped by human rating.



**WILLIAM WILKINSON'S  
“AN ACCOUNT OF THE PRINCIPALITIES OF  
WALLACHIA AND MOLDOVIA”  
INSPIRED THIS AUTHOR'S  
MOST FAMOUS NOVEL**



**“Who was Bram Stoker?”**

All 3 contestants got this final question right, but Watson won.  
Former champion Jennings said:  
“I for one welcome our new computer overlords.”

# **LANGUAGE AND THOUGHT**

# Language and Thought

Things in the world



Thoughts

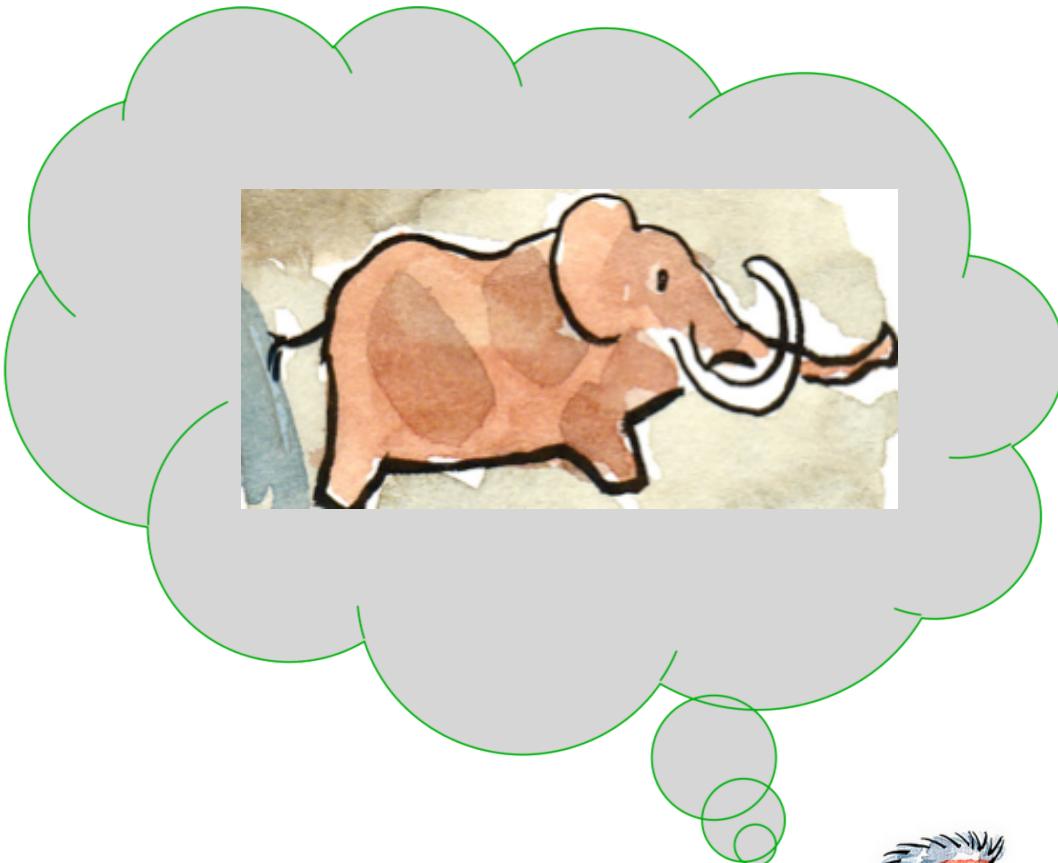
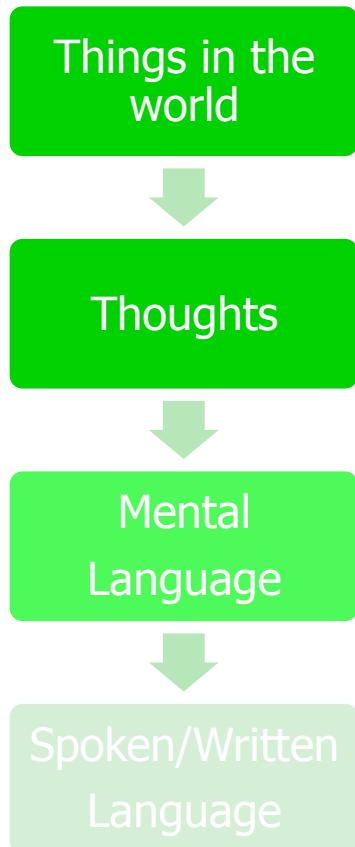


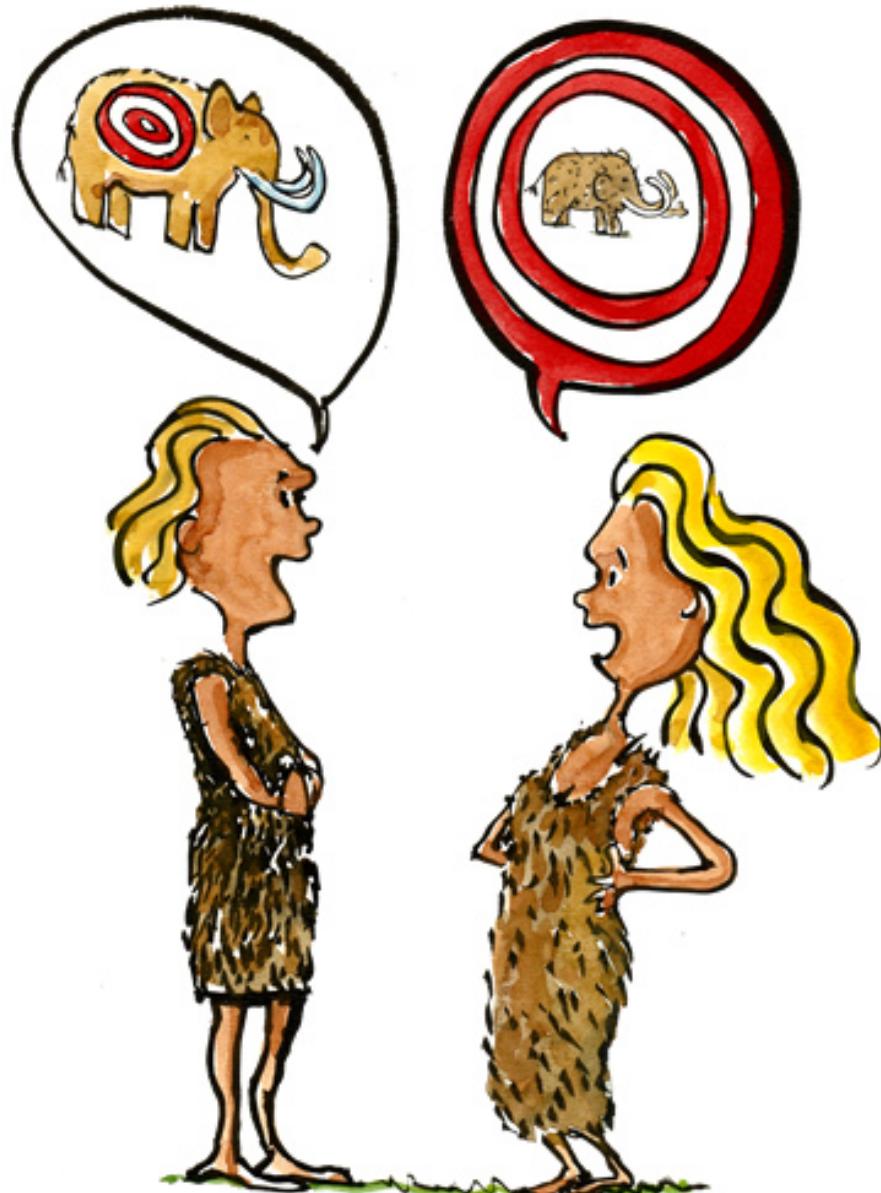
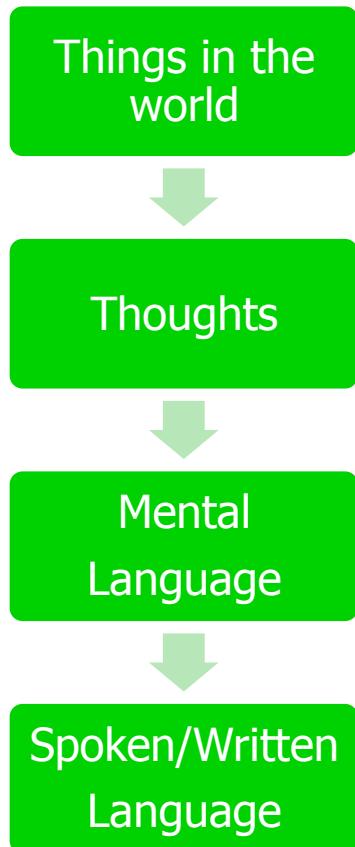
Mental Language



Spoken/Written Language







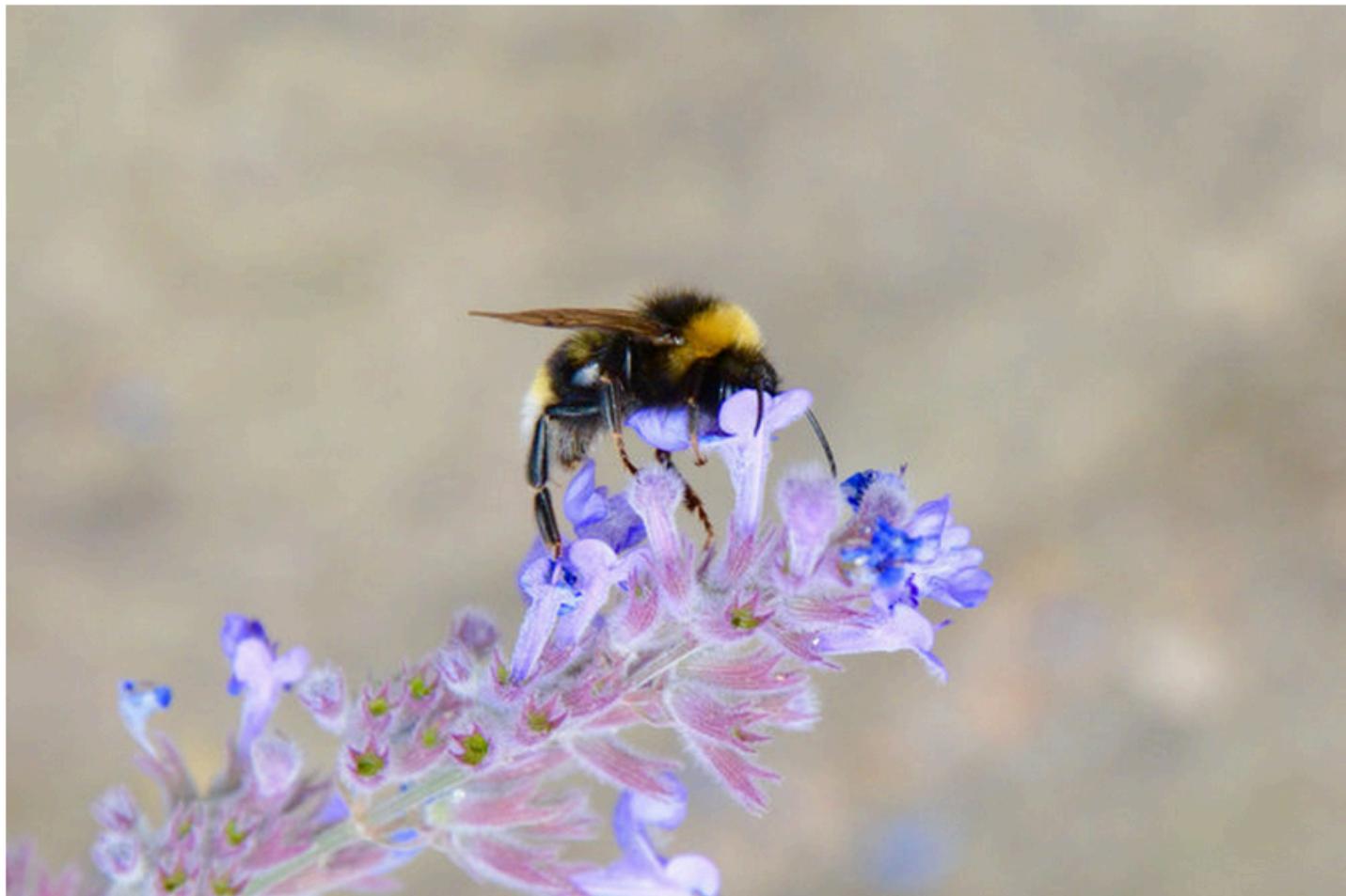
# Much of Meaning is Below the Surface of Words



## SCIENCE

# *Climate Change Is Shrinking Where Bumblebees Range*

By NICHOLAS ST. FLEUR JULY 9, 2015



A buff-tailed bumblebee on catmint flowers. Warming temperatures have caused bumblebee populations to retreat from the southern limits of their ranges in North America and Europe, a new study found. *James T.K. Hager*

(1 min solo, 3 mins in conversation with your neighbor)

## **WHY IS NLP DIFFICULT?**

# Why is NLP difficult?

- Computers are not brains
  - There is evidence that much of language understanding is built-in to the human brain
- Computers do not socialize
  - Much of language is about communicating with people
- Key problems:
  - Representation of *meaning*
  - Language presupposes knowledge about the world
  - Language reflects the *surface* of meaning
  - Language *presupposes* communication between people.

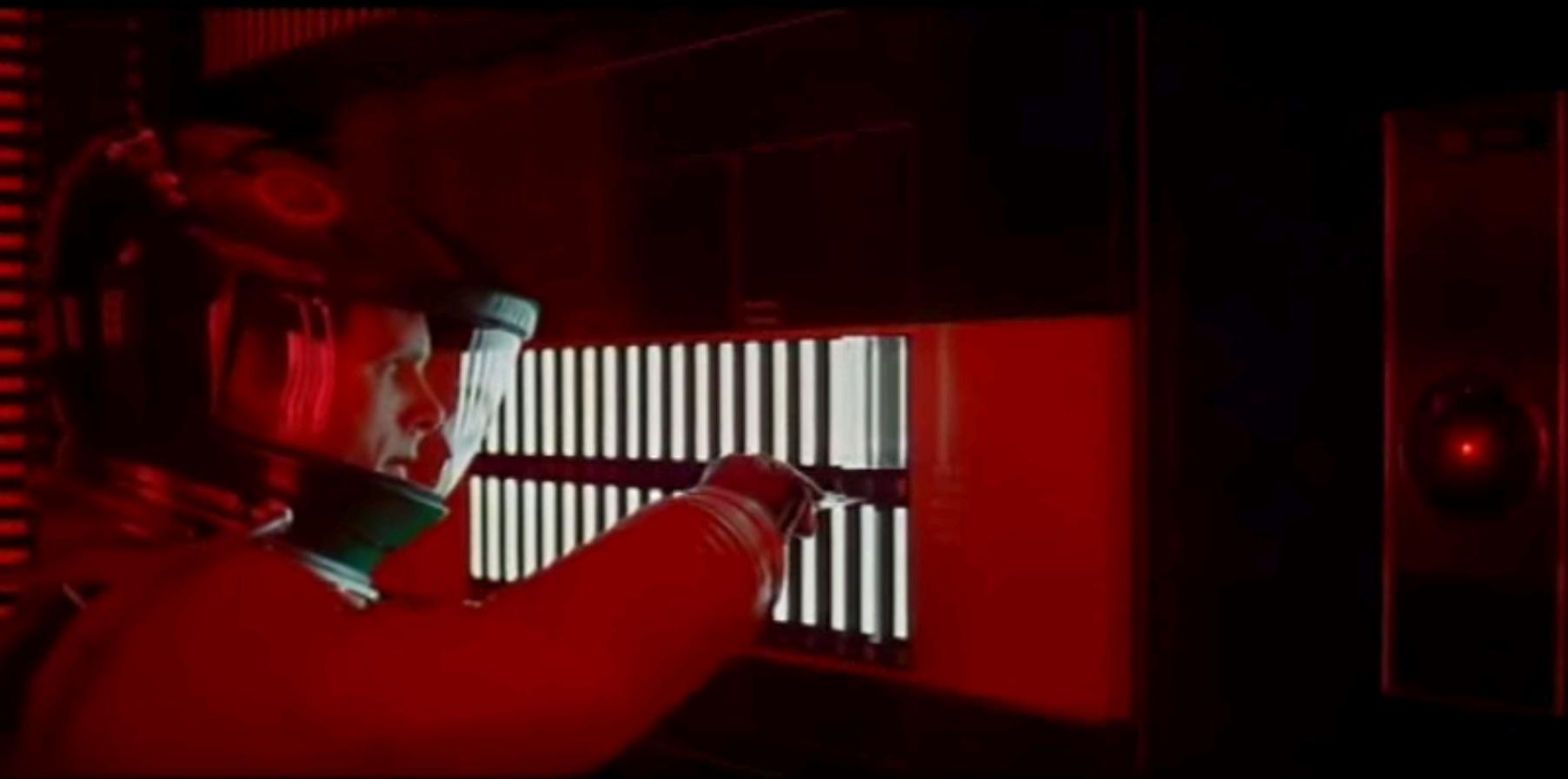


**WHAT CAN'T WE DO YET?**



0:11 / 2:31





1:27 / 4:35



# Course Details

- **Readings:**
  - Free online: <http://nltk.org/book>
  - Other readings will be assigned.
  - Reference: Jurafsky and Martin, *Speech and Language Processing*, 2nd Edition (not 1<sup>st</sup>)
- **Prerequisites:**
  - I206 or equivalent
  - Python programming (200 line program)
  - Basic statistics
- **Work and Grading:**
  - Lots of in-class and out-of-class work
  - 3 “regular” assignments (individual, group)
  - One final project (group)

# What We'll Do in this Course

- Get the Big Picture
- Learn Some Techniques
  - Code in some cases – get agile!
  - Analysis in others
- Build up tools with shorter assignments
- Learn from each other!
- Final summative project

# Intended Learning Outcomes

(informed by your questionnaires)

- Learn Foundations of Natural Language Processing:
  - What are the major issues
  - What are the standard solutions:
    - How **well** do they work?
    - **How** do they work? (a first introduction)
- Skills:
  - Write NLP programs
    - Quick scripts
    - End-to-end solution for a well-sscoped problem
  - Assess NLP problems
    - Know which approaches to apply when
    - Understand the difficulties
    - Be skeptical about over-inflated claims!

# Active + Peer Learning

- Change lectures from passive to active, by:
  - Pausing frequently
  - Asking students to reflect on and build on concepts
- BOTH individual AND collaborative work
  - First think individually
  - Then interact with other student(s)
    - Individual accountability AND
    - Positive interdependence
- Requires student and instructor preparation
  - MUST do work before class to prepare
  - In-class work is limited to short exercises

# What we are NOT Covering

- We are focusing on the NLP parts that you don't get in other classes, so take other classes to learn about:
  - Advanced Machine Learning Algorithms
    - Details of “Deep Learning”
  - Algorithms at Scale (on Spark, Hadoop, etc)
  - Recommender Systems
- Within NLP we are NOT covering:
  - Machine Translation
  - Many other topics! It's a big field!

# Software

- bCourses – be sure notifications are set to send announcements at least daily (not weekly).
- Come to class with your laptop and prepared to code
  - Python 3.X
  - NLTK 3.0
  - ipython Notebook
  - Anaconda environment is fantastic!

# Course Ground Rules

- Attend every class
- Be on time (Berkeley time)
- Code Code of Conduct
  - Do not copy code without attribution
  - The same goes for other work; you can talk with others but you must do your own work unless instructed otherwise.
  - Do not misrepresent results.
  - Violating the Berkeley code of conduct is grounds for failing the course.

# Assignments for Mon and Wed

- Details will be posted on bCourses by today midnight
- By Monday class:
  - Read up on ipython and ipython notebook
  - Practice work with ipython notebook
  - Think about a text collection to “adopt”
- By Wednesday class:
  - Read NLTK book Sections 1.1-1.3
  - Practice work based on these readings