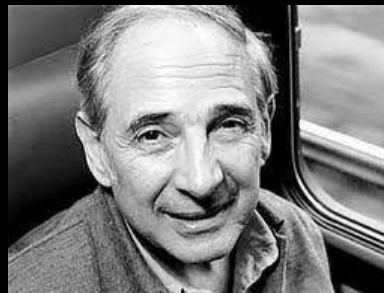


"Natural language is the most important part of artificial intelligence."

John Searle



"Natural language processing is a cornerstone of artificial intelligence, allowing computers to read and understand human language, as well as to produce and recognize speech."

Ginni Rometty



"Natural language processing is one of the most important fields in artificial intelligence and also one of the most difficult."

Dan Jurafsky



What is Natural Language Processing (NLP)?

Natural language processing is the set of methods for making human language accessible to computers

(Jacob Eisenstein)



Natural language processing is the field at the intersection of Computer science (Artificial intelligence) and linguistics

(Christopher Manning)



Make computers to understand natural language to do certain task humans can do such as
Machine translation, Summarization, Questions answering

(Behrooz Mansouri)



Example: Conversational Agent

Conversational agents contain:

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

David Bowman:

Open the pod bay doors, Hal.

HAL:

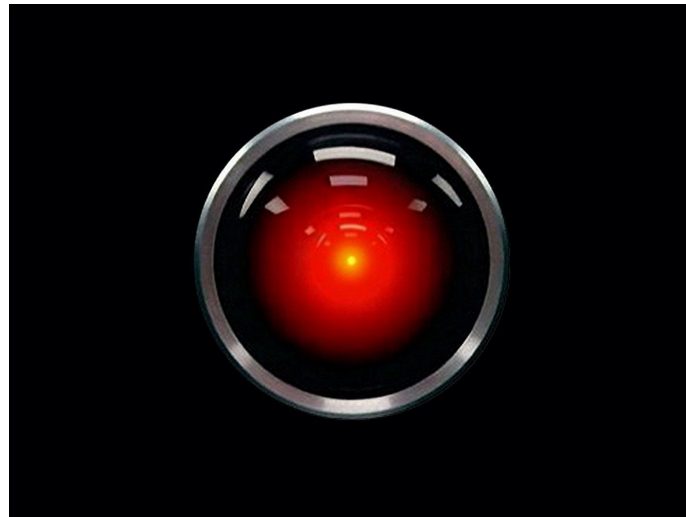
I'm sorry, Dave, I'm afraid I can't do that.

David Bowman:

What are you talking about, Hal?

...HAL:

I know that you and Frank were planning to disconnect me, and I'm afraid that's something I cannot allow to happen.



2001: A Space Odyssey – [HAL 9000](#)

HAL is an artificial agent capable of such advanced language-processing behavior as speaking and understanding English, and at a crucial moment in the plot, even reading lips

Natural Language Processing: Terms

Natural language refers to the language that humans use to communicate with each other, such as English, Spanish, or Chinese

Processing

As distinguished from data processing

Question: How is data processing and natural language processing different?

Natural Language Processing: Terms

Consider the Unix `wc` program, which counts the total number of bytes, words, and lines in a text file

- When used to count bytes and lines, `wc` is an ordinary **data processing** application
- However, when it is used to count the words in a file, it requires **knowledge** about what it means to be a word and thus becomes a **language processing** system

Natural Language Processing vs Computational Linguistics

In **linguistics**, language is the object of study

- Computational methods may be brought to bear, just as in scientific disciplines like computational biology and computational astronomy, but they play only a supporting role

In contrast, **natural language processing** is focused on the design and analysis of computational algorithms and representations for processing natural human language

- The goal of natural language processing is to provide new computational capabilities around human language: for example, extracting information from texts, translating between languages, answering questions, holding a conversation, taking instructions

Knowledge Requirement for Machine

Machines require much broader and deeper knowledge of language

What does HAL need?

- Recognize words from an audio signal and to generate an audio signal from a sequence of words
 - knowledge about **phonetics** and phonology: how words are pronounced in terms of sequences of sounds
- HAL is capable of producing contractions like *I'm* and *can't*
 - knowledge about **morphology**, the way words break down into component parts that carry meanings
- HAL must use structural knowledge to properly string together the words that constitute its response
 - knowledge needed to order and group words comes under the heading of **syntax**
- ...