

CMPT 825

Natural Language Processing

Angel Xuan Chang

angelxuanchang.github.io/nlp-class

Programming Languages

C, C++, Java, Python, ...

- unambiguous
- fixed
- designed
- learnable?
- known simple semantics

Natural Languages

French, English, Korean, Chinese, Tagalog, ...

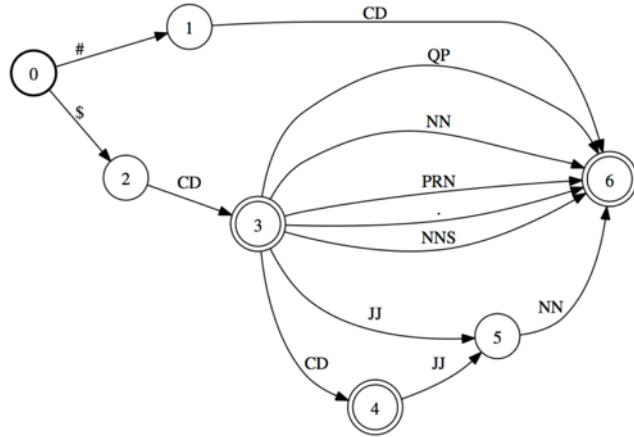
- ambiguous
- evolving
- transmitted
- learnable
- complex semantics

Why is NLP computationally hard?

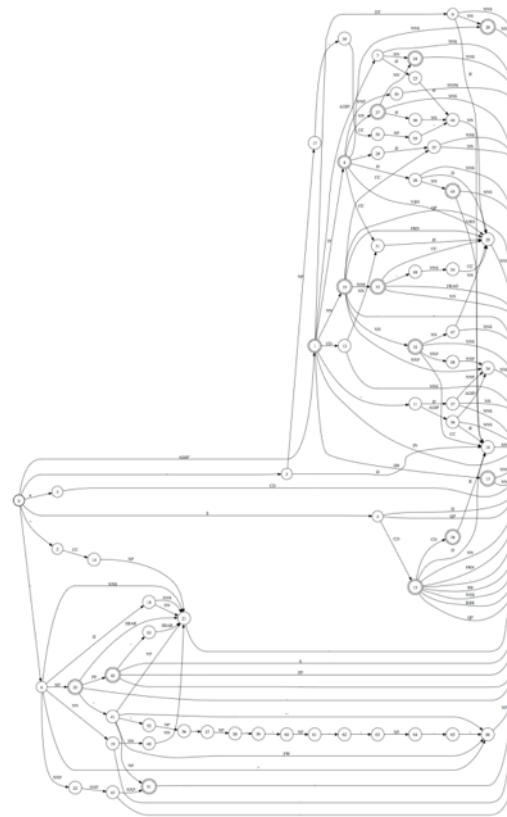
\$ 3 monthly fee
\$ 1 pershare pricing
\$ CD JJ NN a noun phrase type

10 noun phrase types

100 noun phrase types



Language
is complex



Language is diverse

- Many ways of saying the same thing
 - “Siri, play some music for us”,
 - “Siri, please play a song”,
 - ...
- Both at word and sentence level
 - “music” vs “song” vs “tune”
 - “play something nice”

Language requires shared context

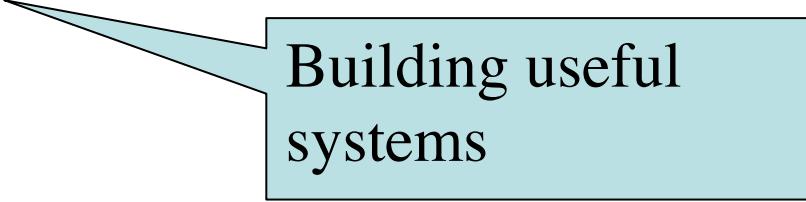
- “Play it again”
- “Can you put the dishes away?”
 what dishes?
 away where?
- “Let’s meet by the bridge at 4”

Language is ambiguous

- Lung cancer in women mushrooms
 - Mushrooms is noun or a verb?
- Ban on nude dancing on governor's desk
 - Similar to “if-then-else” ambiguity
- Island Monks Fly in Satellite to Watch Pope Funeral
 - “**fly in**” vs. “**fly [OBJ in Satellite]**” hidden segmentation
- British Left Waffles on Falkland Islands
 - Is it **British/Noun Left/Verb** or **British Left/NP Waffles/Verb**?

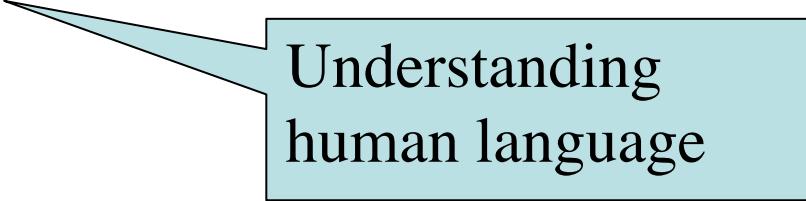
What is the difference?

NLP = Natural Language Processing



Building useful
systems

CL = Computational Linguistics



Understanding
human language

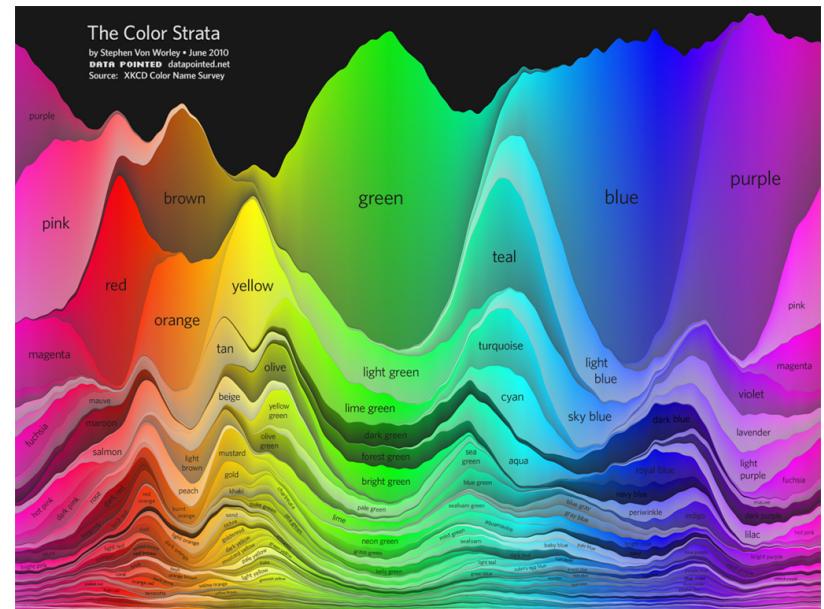
What is language?

- Language is used to communicate
 - Things, actions, abstract concepts



What is language?

- Language puts categories on the world
 - It discretizes a continuous space



What is language?

- Language picks out what is salient and important
- What concepts do we have words for?
- Different languages have different discretization boundaries



Different levels of language

- **Phonetics:** acoustic and perceptual elements
- **Phonology:** basic sounds (phonemes) and rules for combination
 - e.g. syllable constraints in Japanese. E.g. English loanword for “dog” in Japanese is “doggu”
- **Morphology:** how morphemes combine to form words, relationship of phonemes to meaning
 - e.g. delight-ed vs. de-light-ed

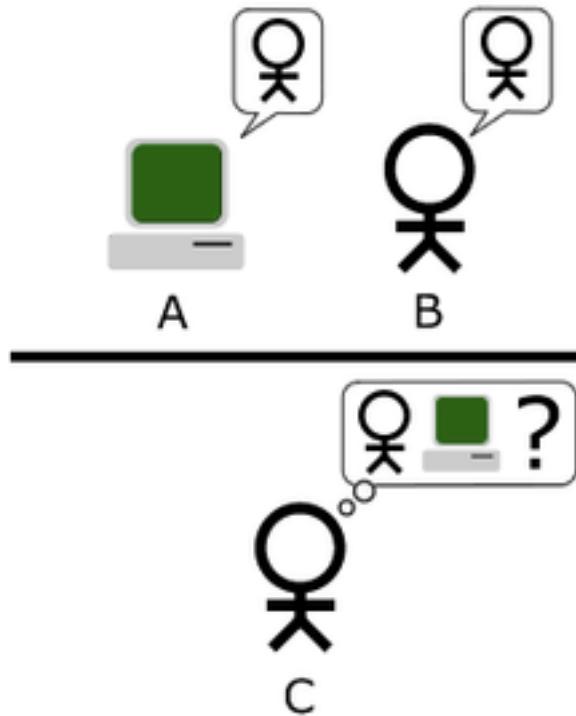
Different levels of language

- **Syntax:** sentence formation
 - e.g. The clown who the musician hits watches the ballerina
- **Semantics:** meaning (from syntax to logical formulas)
 - e.g. Everyone is not here => what does this mean? Nobody / Not everyone is here.
- **Pragmatics:** meaning that is not part of compositional meaning,
 - e.g. “The ham sandwich wants another coffee”

Imagine an "**Imitation Game**," in which a man and a woman go into separate rooms and guests try to tell them apart by writing a series of questions and reading the typewritten answers sent back. In this game both the man and the woman aim to convince the guests that they are the other.



Alan Turing



We now ask the question, "**What will happen when a machine takes the part of A in this game?**" Will the interrogator decide wrongly as often when the game is played like this as he does when the game is played between a man and a woman? These questions replace our original, "**Can machines think?**"

Winograd Schema

The town councillors refused to give the angry demonstrators a permit because they feared violence.

Who feared violence?



Answer 0: the town councillors

Answer 1: the angry demonstrators

The town councillors refused to give the angry demonstrators a permit because they advocated violence.

Who advocated violence?

Answer 0: the town councillors



Answer 1: the angry demonstrators

Some examples of NLP achievements

IBM Watson plays Jeopardy



Speech to Speech Translation



KIT Lecture Translator



NECT Speech Translator



Skype Translator

Topics in NLP research

- Dialogue and Interactive Systems
- Discourse and Pragmatics
- Generation and Summarization
- Information Extraction and Question Answering
- Information Retrieval
- Language Resources and Evaluation
- Language and Vision
- Linguistic and Psycholinguistic Aspects of CL
- Machine Learning for NLP
- Machine Translation
- NLP for Web, Social Media and Social Sciences
- NLP-enabled Technology
- Phonology, Morphology and Word Segmentation
- Semantics
- Sentiment Analysis and Opinion Mining
- Spoken Language Processing
- Tagging, Chunking, Syntax and Parsing
- Text Categorization and Topic Models