Introduction to Natural Language Processing

Introduction

book

https://web.stanford.edu/~jurafsky/slp3/ed3book_dec302020.pdf

Hit record

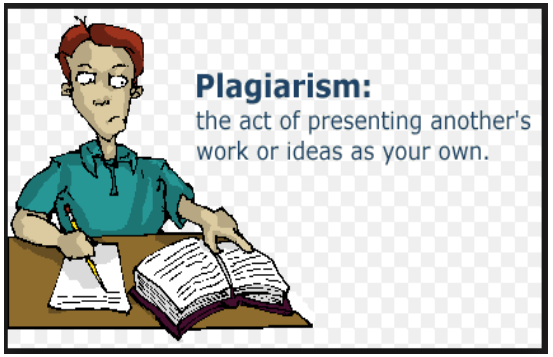
Syllabus

The semester

Don’t plagiarize

Academic Misconduct

On August 12, 2015, this Interim VCU Honor System was approved for use during the 2015/16 academic year; reports made on or after August 19, 2015, will be adjudicated under this policy. Please contact the Office of Student Conduct and Academic Integrity if you have questions regarding a matter that occurred under the prior VCU Honor System or to retrieve a copy of the former policy.



The Honor System prohibits the following acts of academic misconduct as defined below:

Plagiarism: representing the words, ideas, facts, opinions, theories, illustrations, tables or any part of another’s work as one’s own on an academic assignment without customary and proper acknowledgment of the source.

Cheating: receiving, giving, or attempting to receive or give unauthorized assistance, such as materials, devices, information, notes or sources, related to academic matters.

Small Lecture

What is Natural Language Processing?

- The study of human languages and how they can be represented computationally and analyzed and generated algorithmically
  - Understanding:
    - *The cat is on the mat.* --> on (cat, mat)
  - Generation:
    - on (cat, mat) --> *The cat is on the mat.*
    - on (dog, couch) -> *The dog is on the couch.*
- Studying NLP involves studying natural language, formal representations, and algorithms for their manipulation

What *is* Natural Language Processing?

Building computational models of natural language comprehension and production

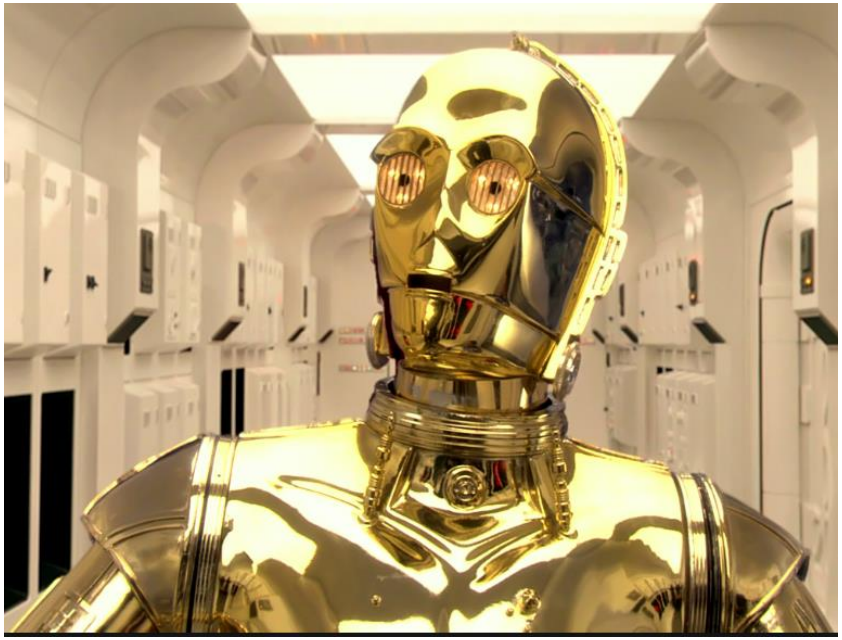
Other Names:

- Computational Linguistics (CL)
- Human Language Technology (HLT)
- Natural Language Engineering (NLE)
- Speech and Text Processing

Perspectives

- Engineering Perspective
- Cognitive Science Perspective
- Theoretical Linguistics Perspective

Engineering Perspective



Use CL as part of a larger application:

- Spoken dialogue systems for telephone based information systems
- Components of web search engines or document retrieval services
  - Machine translation
  - Question/answering systems
  - Text Summarization
- Interface for intelligent tutoring/training systems

Emphasis on

- Robustness (doesn't collapse on unexpected input)
- Coverage (does something useful with most inputs)
- Efficiency (return an answer in a timely manner)

Cognitive Science Perspective

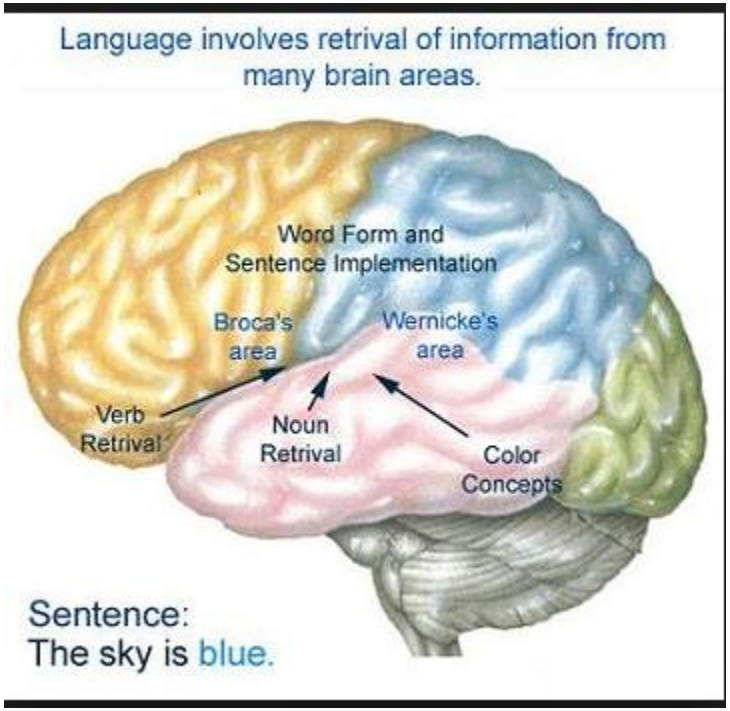
Goal: gain an understanding of how people comprehend and produce language.

Goal: a model that explains actual human behaviour

Solution must:
- explain psycholinguistic data
- be verified by experimentation

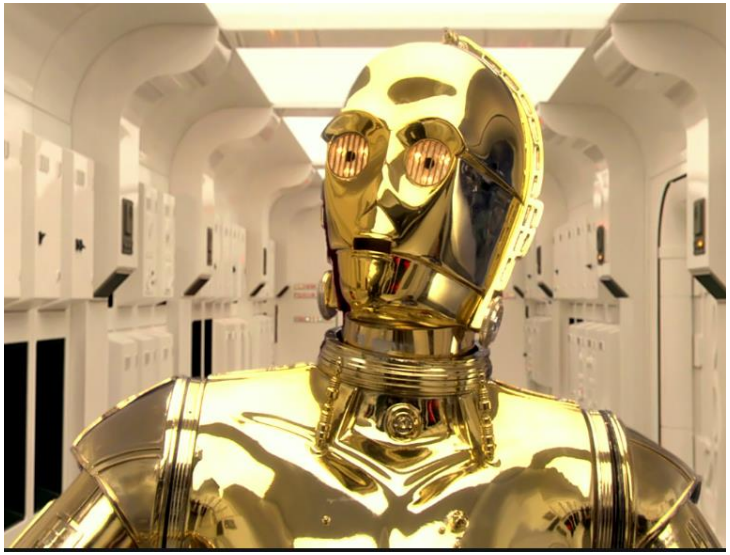
Theoretical Linguistics Perspective

- In principle, coincides with the Cognitive Science Perspective
  - CL can potentially help test the empirical adequacy of theoretical models.
  - Building computational models of the theories allows them to be empirically tested.
    - E.g., does your grammar correctly parse all the grammatical examples in a given test suite, while rejecting all the ungrammatical examples?



Orientation of this Class

- Emphasis on principles and techniques
- Emphasis on processing textual input (as opposed to speech)
- Oriented towards both symbolic and statistical approaches



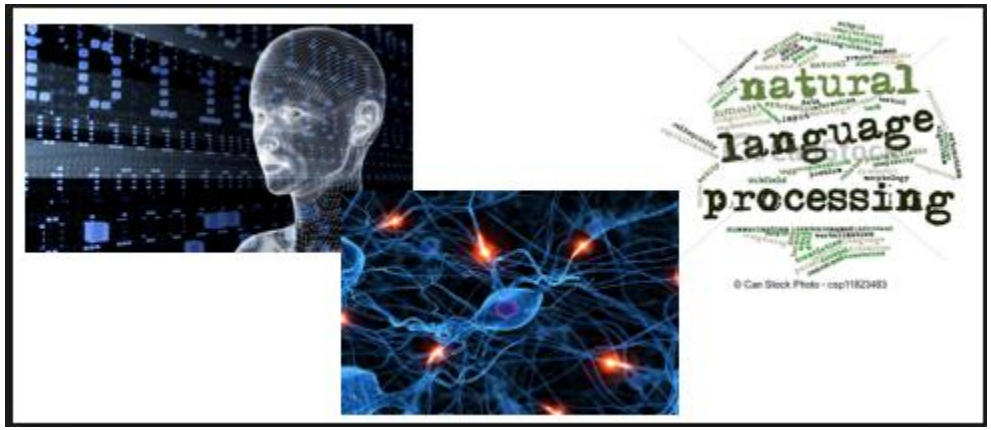
Why Should You Care?

Trends:

1. An enormous amount of knowledge is now available in machine readable form as natural language text

2. Conversational agents are becoming an important form of human-computer communication

3. Much of human-human communication is now mediated by computers



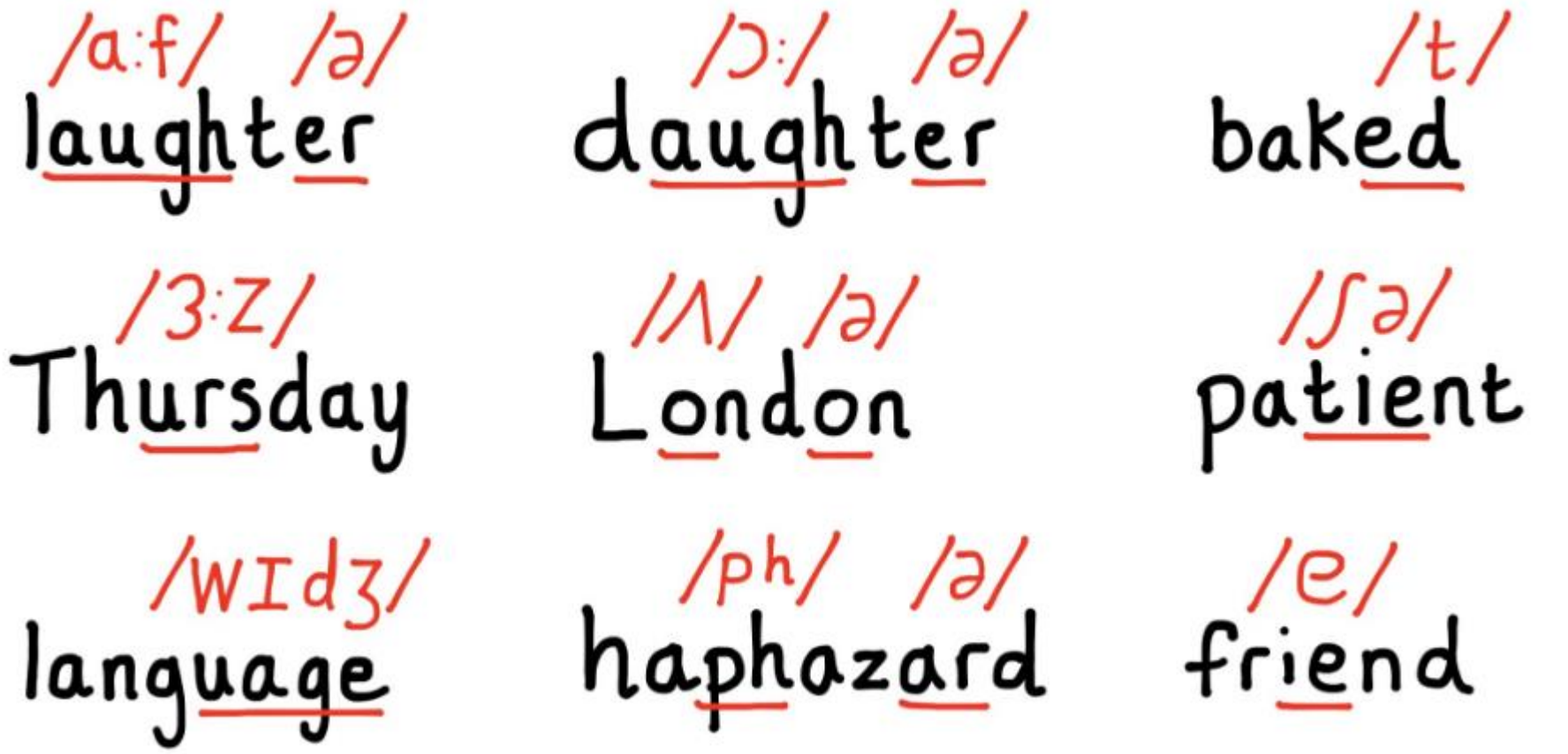
Knowledge needed to understand and produce language

- ***Phonetics and phonology*:**
  - **how words are related to sounds that realize them**
- *Morphology*:
  - how words are constructed from more basic meaning units
- *Syntax*:
  - how words can be put together to form correct utterances
- *Lexical semantics*:
  - what words mean
- *Compositional semantics*:
  - how word meanings combine to form larger meanings
- *Pragmatics*:
  - how situation affects interpretation of utterance
- *Discourse structure*:
  - how preceding utterances affects processing of next utterance

Phonetics and Phonology

speech sounds, their production, and the rule systems that govern their use

---

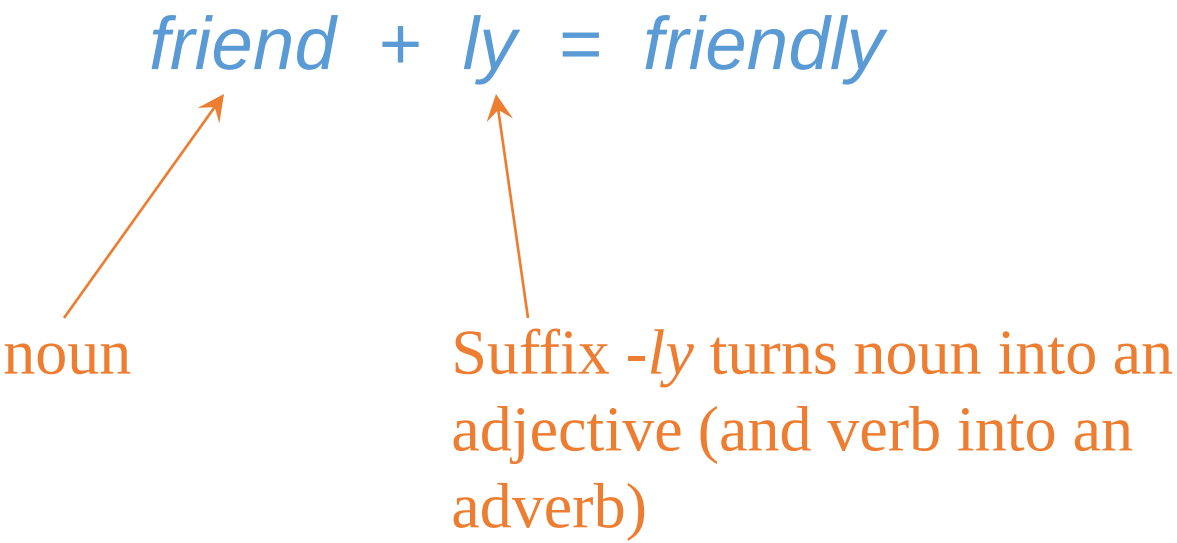


Knowledge needed to understand and produce language

- *Phonetics and phonology:*
  - how words are related to sounds that realize them
- ***Morphology:***
  - **how words are constructed from more basic meaning units**
- *Syntax:*
  - how words can be put together to form correct utterances
- *Lexical semantics:*
  - what words mean
- *Compositional semantics*:
  - how word meanings combine to form larger meanings
- *Pragmatics:*
  - how situation affects interpretation of utterance
- *Discourse structure:*
  - how preceding utterances affects processing of next utterance

Morphology

- How words are constructed from more basic units, called *morphemes*



Morphology

- Words and their composition can be tricky
    - Pluralization:
        - cat -> cats
        - dog -> dogs

and then we have:

- child -> children
- mouse -> mice

mouse

mouses

Morphology

- Words and their composition can be tricky
  - Pluralization:
    - cat -> cats
    - dog -> dogs

and then we have:

  - child -> children
  - mouse -> mice

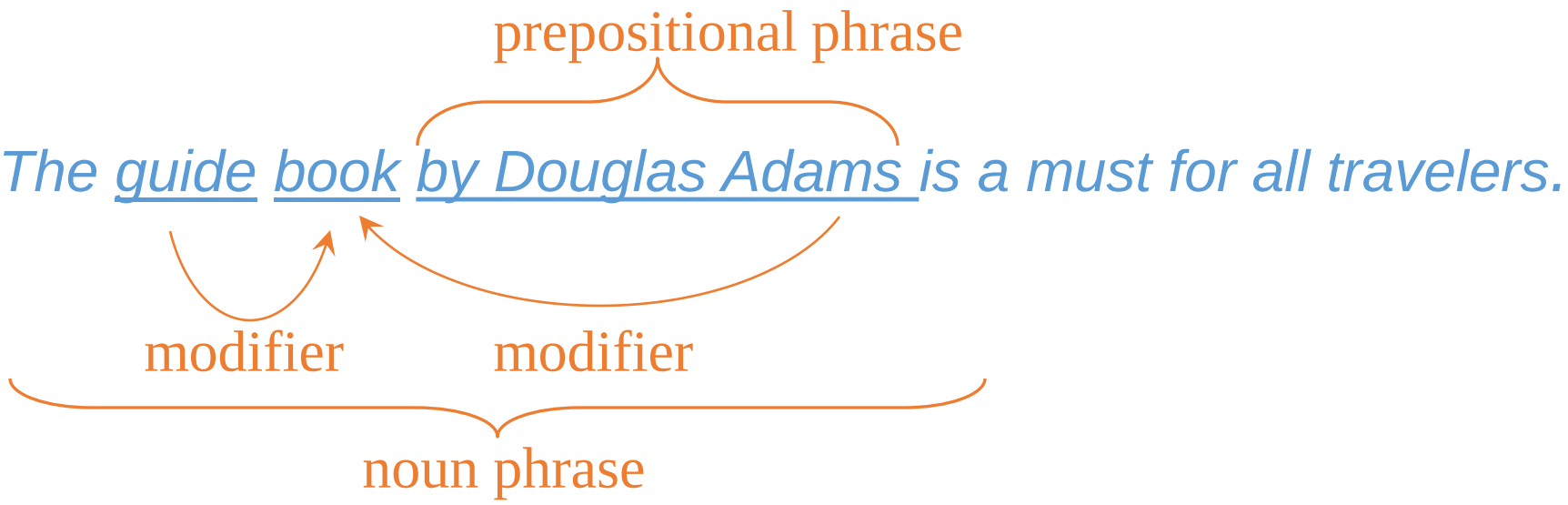
  - Suffixes (or not)
    - undo, union

Knowledge needed to understand and produce language

- *Phonetics and phonology:*
  - how words are related to sounds that realize them
- *Morphology:*
  - how words are constructed from more basic meaning units
- ***Syntax:***
  - **how words can be put together to form correct utterances**
- *Lexical semantics:*
  - what words mean
- *Compositional semantics*:
  - how word meanings combine to form larger meanings
- *Pragmatics:*
  - how situation affects interpretation of utterance
- *Discourse structure:*
  - how preceding utterances affects processing of next utterance

Syntactic Knowledge

- how words can be put together to form legal sentences in the language
- what structural role each word plays in the sentence
- what phrases are subparts of other phrases



- Syntax: the structuring of words into larger phrases
  - Number of ways to structure a phrase and achieve the same meaning
    - John hit Bill (active)
    - Bill was hit by John (passive)
    - Bill, John hit (preposing)
    - Who John hit was Bill (wh-cleft)

Knowledge needed to understand and produce language

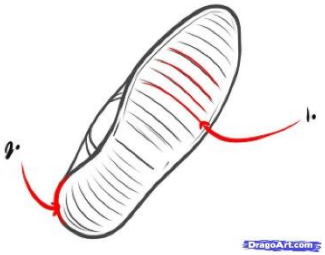
- *Phonetics and phonology:*
  - how words are related to sounds that realize them
- *Morphology:*
  - how words are constructed from more basic meaning units
- *Syntax:*
  - how words can be put together to form correct utterances
- ***Lexical semantics:***
  - **what words mean**
- *Compositional semantics*:
  - how word meanings combine to form larger meanings
- *Pragmatics:*
  - how situation affects interpretation of utterance
- *Discourse structure:*
  - how preceding utterances affects processing of next utterance

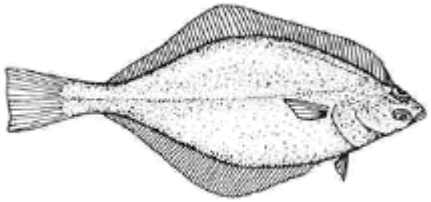
Semantic Knowledge

- What does a word mean?

bat

sole





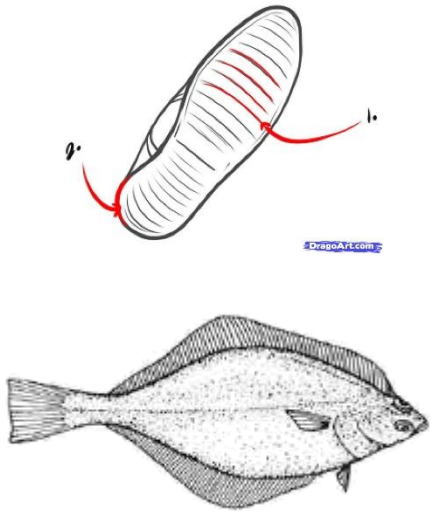
Knowledge needed to understand and produce language

- *Phonetics and phonology:*
  - how words are related to sounds that realize them
- *Morphology:*
  - how words are constructed from more basic meaning units
- *Syntax:*
  - how words can be put together to form correct utterances
- *Lexical semantics:*
  - what words mean
- ***Compositional semantics*:**
  - **how word meanings combine to form larger meanings**
- *Pragmatics:*
  - how situation affects interpretation of utterance
- *Discourse structure:*
  - how preceding utterances affects processing of next utterance

Semantic Knowledge

- What does a word mean?

sole



I ate sole for dinner.

bat

The bat flew.

Knowledge needed to understand and produce language

- *Phonetics and phonology:*
  - how words are related to sounds that realize them
- *Morphology:*
  - how words are constructed from more basic meaning units
- *Syntax:*
  - how words can be put together to form correct utterances
- *Lexical semantics:*
  - what words mean
- *Compositional semantics*:
  - how word meanings combine to form larger meanings
- ***Pragmatics:***
  - **how situation affects interpretation of utterance**
- *Discourse structure:*
  - how preceding utterances affects processing of next utterance

Pragmatics and Discourse: The influence of Context on language

"Going Home" – A play in one act; five scenes

by Bonnie Dorr

- Scene 1: Pennsylvania Station, NY
  - Bonnie: Long Beach?
  - Passerby: Downstairs, LIRR Station.
- Scene 2: Ticket Counter, LIRR Station
  - Bonnie: Long Beach?
  - Clerk: $4.50.

- Scene 3: Information Booth, LIRR Station
  - Bonnie: Long Beach?
  - Clerk: 4:19, Track 17.
- Scene 4: On the train, vicinity of Forest Hills
  - Bonnie: Long Beach?
  - Clerk: Change at Jamaica.
- Scene 5: On the next train, vicinity of Lynbrook
  - Bonnie: Long Beach?
  - Clerk: Right after Island Park.

- Scene 3: Information Booth, LIRR Station
  - Bonnie: Long Beach?
  - Clerk: 4:19, Track 17.
- Scene 4: On the train, vicinity of Forest Hills
  - Bonnie: Long Beach?
  - Clerk: Change at Jamaica.
- Scene 5: On the next train, vicinity of Lynbrook
  - Bonnie: Long Beach?
  - Clerk: Right after Island Park.

1. *Bonnie has a limited vocabulary*
2. *Context is important*

Knowledge needed to understand and produce language

- *Phonetics and phonology:*
  - how words are related to sounds that realize them
- *Morphology:*
  - how words are constructed from more basic meaning units
- *Syntax:*
  - how words can be put together to form correct utterances
- *Lexical semantics:*
  - what words mean
- *Compositional semantics*:
  - how word meanings combine to form larger meanings
- *Pragmatics:*
  - how situation affects interpretation of utterance
- ***Discourse structure:***
  - **how preceding utterances affects processing of next utterance**

Discourse

- Relationships we infer between discourse entities
- Not expressed in either of the propositions, but from their juxtaposition

Friend 1: *I'm hungry.*

Friend 2: *Let's go to the Fuji Gardens.*

Discourse

- Relationships we infer between discourse entities
- Not expressed in either of the propositions, but from their juxtaposition

Friend 1: *It's a beautiful day.*

Friend 2: *Let's go to the Fuji Gardens.*

Discourse and Temporal Interpretation

Max fell. John pushed him.

explanation

Syntax and semantics: "him" refers to Max

Lexical semantics and discourse: the pushing occurred before the falling.

Knowledge needed to understand and produce language

- *Phonetics and phonology:*
  - how words are related to sounds that realize them
- *Morphology:*
  - how words are constructed from more basic meaning units
- *Syntax:*
  - how words can be put together to form correct utterances
- *Lexical semantics:*
  - what words mean
- *Compositional semantics*:
  - how word meanings combine to form larger meanings
- *Pragmatics:*
  - how situation affects interpretation of utterance
- *Discourse structure:*
  - how preceding utterances affects processing of next utterance

World knowledge

- What we know about the world and what we can assume our hearer knows about the world is intimately tied to our ability to use language

I took the cake from the plate and ate it.

World knowledge

- What we know about the world and what we can assume our hearer knows about the world is intimately tied to our ability to use language

I took the cake from the plate and ate it.

World knowledge

- What we know about the world and what we can assume our hearer knows about the world is intimately tied to our ability to use language

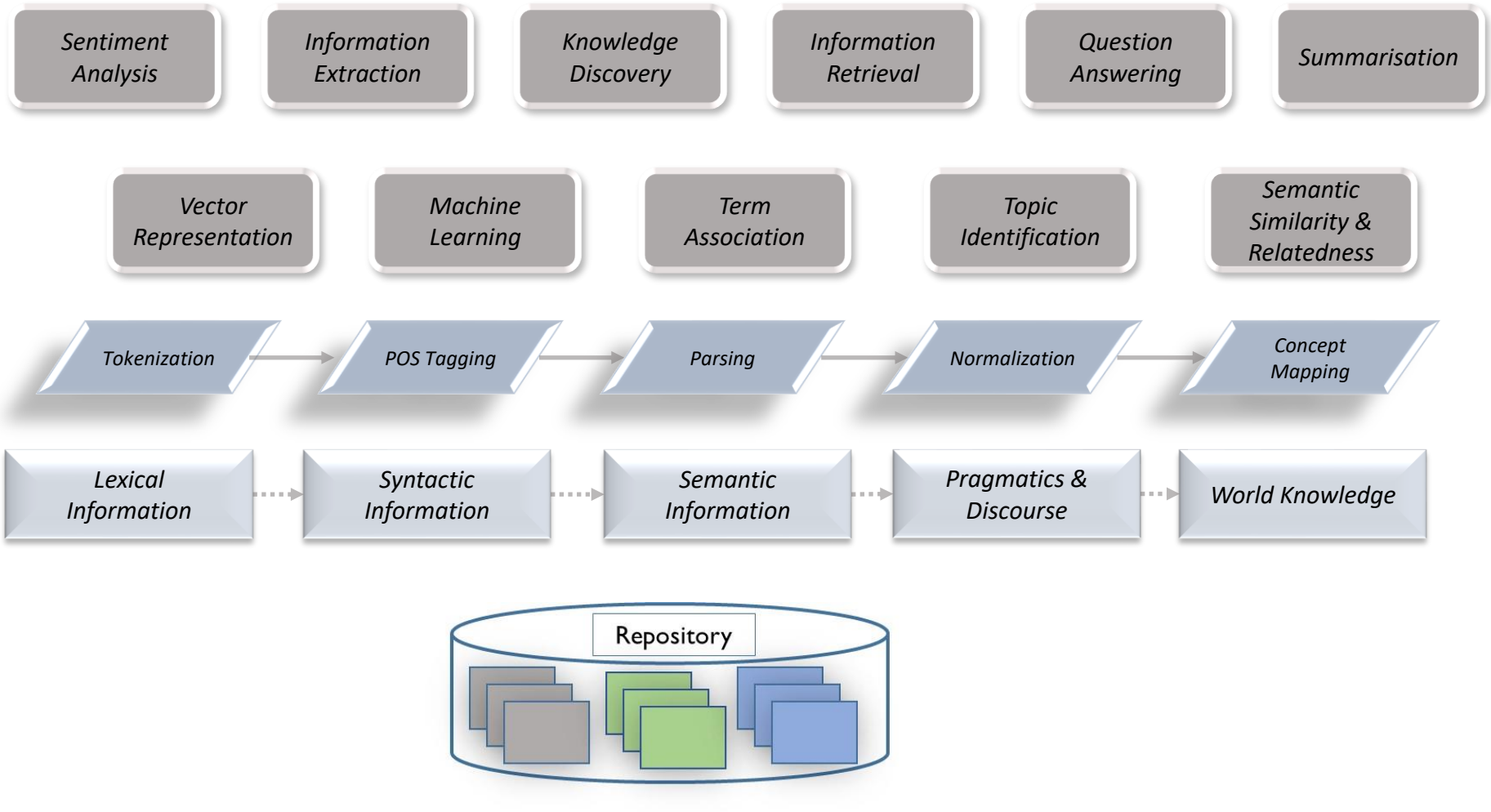
I took the fugu from the plate and ate it.

World knowledge

- What we know about the world and what we can assume our hearer knows about the world is intimately tied to our ability to use language

I took the fugu from the plate and ate it.

Natural language processing stack



Questions?