#### Introduction to CL & NLP

CMSC 35100 April 1, 2003

## Speech and Language Processing

- Language applications
  - Language understanding, Question-answering,
    Information extraction, Speech recognition, Machine
    Translation,...
- Computational Linguistics
  - Modeling language structure
  - Modeling human use of language
    - What does it mean to "know" a language?

# Models and Methods from Many Fields

- Linguistics:Morphology, phonology, syntax, semantics,..
- Psychology:Reasoning, mental representations
- Formal logic
- Philosophy (of language)
- Theory of Computation: Automata,..
- Artificial Intelligence: Search, Reasoning, Knowledge representation, Machine learning, Pattern matching
- Probability..

## Balancing Act

- Competitive & integrative approaches:
  - Symbolic vs Stochastic
- Early approaches: 40's & 50's
  - Formal language theory (Chomsky, Backus)
    - Automata theory
  - Probabilistic techniques (Shannon):
    - Noisy channel model
    - Decoding

#### Two Paths: '50-'83

- Symbolic:
  - Formal language theory (Chomsky, Harris)
  - Logic-based systems (Kaplan,Kay)
    - Lexical functional grammar, feature systems
  - Toy symbolic NLU systems: (Winograd, Woods,)
    - Blocks world, Lunar, ..
  - Discourse modeling: (Grosz, Sidner, Webber)
    - Reference, Topic and Task structure
- Stochastic: (Jelinek, Brown, Baker, Bahl, Rabiner)

# To the Present: Empiricism & Moore's Law

- Empiricism:
  - Finite State methods: (Kaplan&Kay, Church)
    - Morphology, Syntax, .
  - Probabilistic approaches (Jelinek, Perreira, Charniak)
    - Tagging, syntax, parsing, discourse,...
- Moore's Law:
  - Data-driven (and probabilistic) techniques demand processor speed, disk space, memory!!

## Language & Intelligence

- Turing Test: (1949) Operationalize intelligence
  - Two contestants: human, computer
  - Judge: humans
  - Test: Interact via text questions
  - Questions: Which is human???

• Crucially requires language use and understanding

### Limitations of the TuringTest

- ELIZA (Weizenbaum 1966)
  - Simulates Rogerian therapist
    - User: You are like my father in some ways
    - ELIZA: WHAT RESEMBLANCE DO YOU SEE
    - User: You are not very aggressive
    - ELIZA: WHAT MAKES YOU THINK I AM NOT AGGRESSIVE...
  - Passes the Turing Test!! (sort of)
  - "You can fool some of the people...."
- Simple pattern matching technique

### Real Language Understanding

- Requires more than just pattern matching
- But what?,

- **.** 2001:
- Dave: Open the pod bay doors, HAL.
- HAL: I'm sorry, Dave. I'm afraid I can't do that.

#### Phonetics and Phonology

- Convert an acoustic sequence to word sequence
- Need to know:
  - Phonemes: Sound inventory for a language
  - Vocabulary: Word inventory pronunciations
  - Pronunciation variation:
    - Colloquial, fast, slow, accented, context

## Morphology

• Recognitize and produce variations in word forms

- (E.g.) Inflectional morphology:
  - e.g. Singular vs plural; verb person/tense
    - Door + sg: door
    - Door + plural: doors
    - $Be + 1^{st}$  person, sg, present: am

### Syntax

• Order and group words together in sentence

• Open the pod bay doors

– Vs

• Pod the open doors bay

#### Semantics

• Understand word meanings and combine meanings in larger units

- Lexical semantics:
  - Bay: partially enclosed body of water; storage area
- Compositional sematics:
  - "pod bay doors":
    - Doors allowing access to bay where pods are kept

#### Discourse & Pragmatics

• Interpret utterances in context

- Resolve references:
  - "I'm afraid I can't do that"
    - "that" = "open the pod bay doors"
- Speech act interpretation:
  - "Open the pod bay doors"
    - Command



# Ambiguity:

# Language Processing Components

- "I made her duck"
- Means....
  - I caused her to duck down
  - I made the (carved) duck she has
  - I cooked duck for her
  - I cooked the duck she owned
  - I magically turned her into a duck

## Part-of-Speech Tagging

- Ambiguity:
  - Her: pronoun vs possessive adjective
  - Duck: verb vs noun

#### Word Sense Disambiguation

• Ambiguity:

- Make = cook
  - Vs
- Make = carve

#### Syntactic Disambiguation

• I made her duck.



Resources for NLP Systems

- Dictionary
- Morphology and Spelling Rules
- Grammar Rules
- Semantic Interpretation Rules
- Discourse Interpretation

Natural Language processing involves (1) learning or fashioning the rules for each component, (2) embedding the rules in the relevant automaton, (3) and using the automaton to efficiently process the input.