Natural Language Processing

Berlin Chen 2004

Textbooks & References

- ## Textbooks
  - C. Manning and H. Schutze, Foundations of Statistical Natural Language Processing, MIT Press, 1999
  - D. Jurafsky and J. H. Martin, Speech and Language Processing, Prentice-Hall, 2000

- ## References
  - J. Allen, Natural Language Understanding, Benjamin/Cummings Publishing Co, 1995
  - X. Huang, A. Acero, H. Hon, Spoken Language Processing, Prentice Hall, 2001

Motivation for NLP

- **Academic**: Explore the natural of linguistic communication
  - Obtain a better understanding of how language work

- **Practical**: Enable effective human-machine communication
  - Conversational agents are becoming an important form of human-computer communication
  - Revolutionize the way computers are used
    - More flexible and intelligent

Motivation for NLP

- Different Academic Disciplines: Problems and Methods
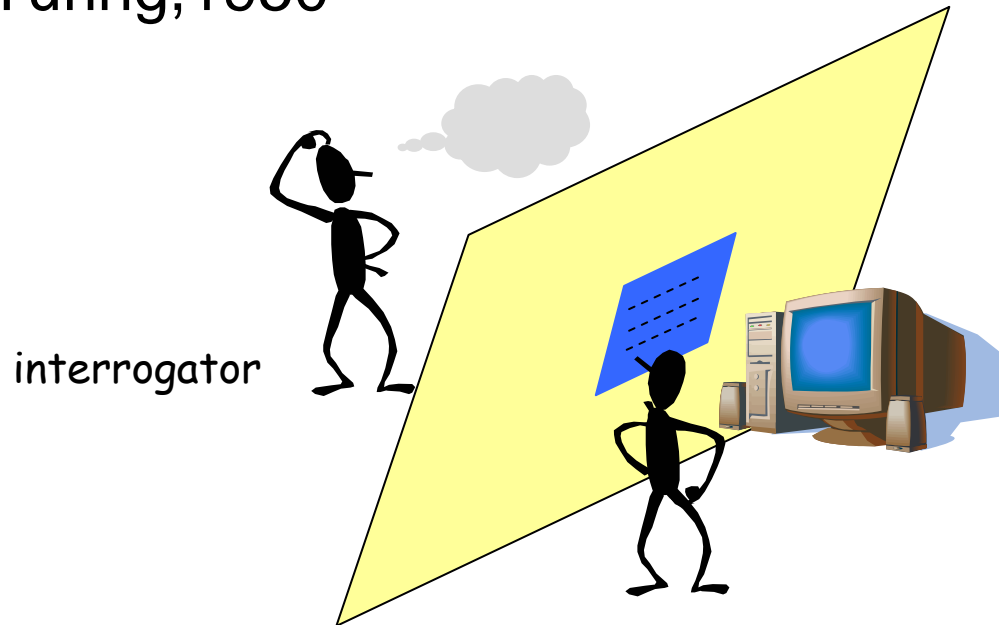 - Electrical Engineering, Statistics
  - Computer Science
  - Linguistics
  - Psychology
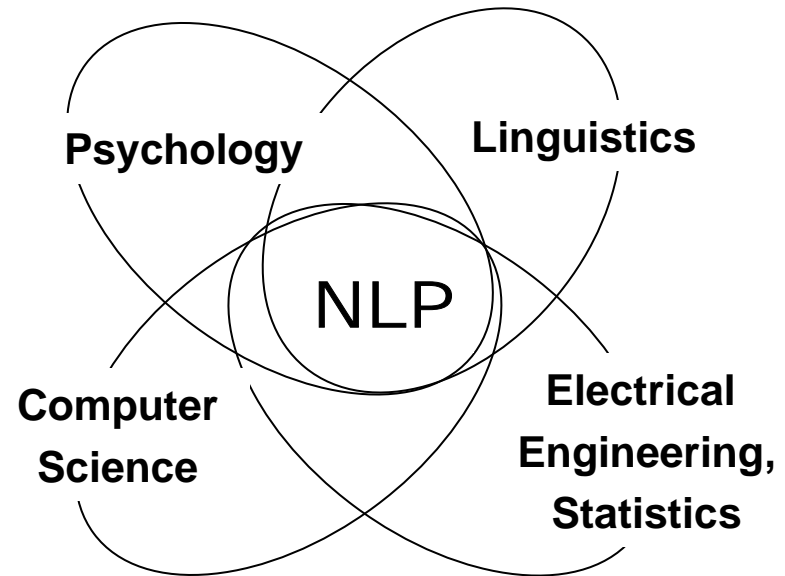
- Many of the techniques presented were first develpoed for speech and then spread over into NLP
  - E.g. Language models in speech recognition

Turing Test

- Alan Turing,1950

interrogator

- – Predicted at the end of 20 century a machine with 10 gigabytes of memory would have 30% chance of fooling a human interrogator after 5 minutes of questions
  - Does it come true?

Hollywood Cinema

- Computers/robots can listen, speak, and answer our questions
  - E.g.: HAL 9000 computer in "*2001: A Space Odyssey*"

    (2001太空漫遊)

State of the Art

- Canadian computer program accepted daily weather data and generated weather reports (1976)

- MIT Spoken dialogue systems for information of restaurant, air travel, etc. (1991~)

- AT&T, How May I Help You?

- Read student essays and grade them
- Automated reading tutor
- ……

State of the Art
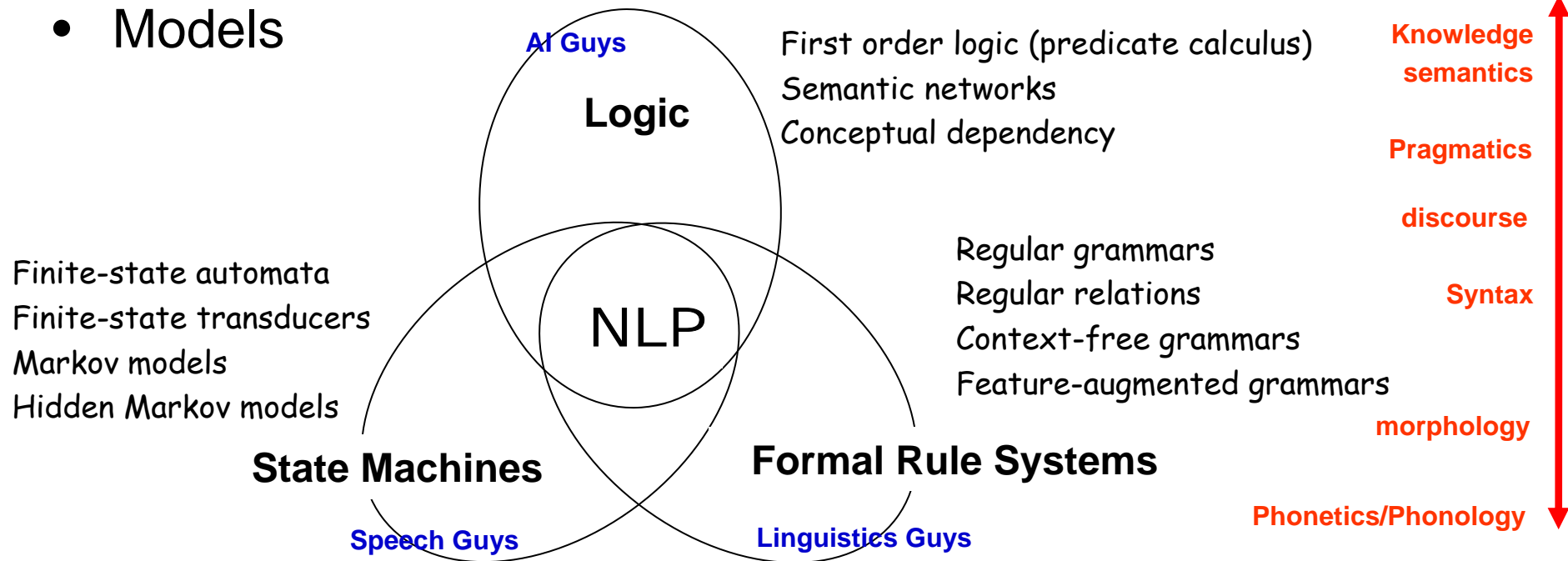
- CMU Universal Speech Interface

Statistical and Non-Statistical NLP

- The dividing line between the two has become much more fuzzy recently
  - An increasing number of non-statistical researches use corpus evidence and incorporate quantitative methods
    - Corpus: "a body of texts" (大量的文稿)

  - Statistical NLP needs to start with all the scientific knowledge available about a phenomenon when building a probabilistic model, rather than closing one's eye and taking a clean-slate approach

Models and Algorithms for NLP

- ## Models

AI Guys

Logic

First order logic (predicate calculus)
Semantic networks
Conceptual dependency

Finite-state automata
Finite-state transducers
Markov models
Hidden Markov models

NLP

Regular grammars
Regular relations
Context-free grammars
Feature-augmented grammars

State Machines

Formal Rule Systems

Speech Guys

Linguistics Guys

Knowledge
semantics

Pragmatics

discourse

Syntax

morphology

Phonetics/Phonology

- ## Algorithms
  - Search:
    - Dynamic programming, depth-first search, best-first search, A* search
  - Learning/Training Methods

Major Topics for NLP

- **Probability Theory/Statistics**
  - Supervised/Unsupervised Machine Learning Techniques

- **Words**
  - Morphology
  - Regular expressions
  - Automata, Finite-State Transducers

- **Syntax**
  - Part-of-Speech Tagging
  - (Probabilistic) Context-Free Grammar
  - Parsing

Major Topics for NLP

- **Semantics/Meaning**
  - Representation of Meaning
  - Semantic Analysis
  - Word Sense Disambiguation

- **Pragmatics**
  - Natural Language Generation
  - Discourse, Dialogue and Conversational Agents
  - Machine Translation

Topic List and Schedule

2/20	Course Overview & Introduction Linguistic Essentials	
2/27	Linguistic Essentials Regular Expressions and Automata	
3/5	Mathematical Foundations 朱惠銘	
3/12	Part-of-Speech Tagging	
3/19	**Break ?** (ICDAT 2004)	
3/26	Collections 張志豪 Parsing with Context-Free Grammars	
4/2	N-gram Language Modeling 黄耀民	
4/9	**Break**	
4/16	Word Sense Disambiguation 劉成韋	
4/23	Midterm	
4/30	Text Categorization 嚴永泰	
5/7	Probabilistic Context-Free Grammars	
5/14	Paper Survey	
5/21	**Break** (ICASSP 2004)	
5/28	Paper Survey	
6/4	Semantics and Logical Form	
6/11	Statistical Alignment and Machine Translation	
6/18	**FINAL**	

Applications of NLP

- Speech Recognition
- Information Retrieval and Extraction
- Summarization
- Question Answering
- Conversational Agents
- Machine (Speech/Language) Translation
- Spelling Check
- Segmentation and Alignment
- Bioinformatics
- ….

Resources

- Corpora (Speech/Language resources)
  - Refer speech waveforms, machine-readable text, dictionaries, thesauri as well as tools for processing them
    - LDC - Linguistic Data Consortium
    - The Association for Computational Linguistics and Chinese Language Processing

Resources

- Institutes/People
  - Foreign
    - MIT
    - CU
    - CMU
    - JHU
    - UMass
    - Cambridge
    - Microsoft
    - IBM
    - MITRE
    - HP
    - ……………

Resources

- Conferences and Journals
  - **ACL**: Association for Computational Linguistics
  - **COLING**: International Conference on Computational Linguistics
  - **Computational Linguistics**
  - **Natural Language Engineering**

  - **ICSLP**: International Conference on Spoken Language Processing
  - **EUROSPEECH:** European Conference on Speech Communication and Technology
  - **ICASSP:** IEEE International Conference on Acoustics, Speech, Signal processing
  - **Speech Communication**
  - **Computer Speech and Language**
  - **IEEE Transactions on Speech and Audio Processing**