Natural Language Processing

Berlin Chen 2003

Textbook & References

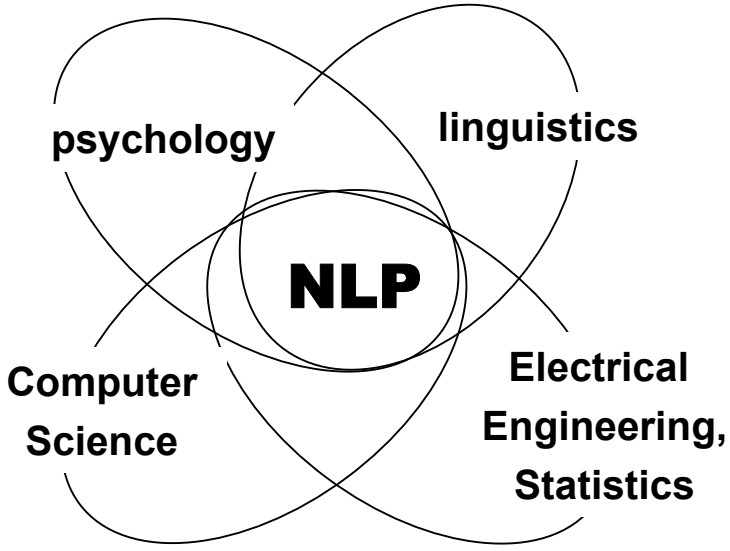
- Textbook
  - D. Jurafsky and J. H. Martin, Speech and Language Processing, Prentice-Hall, 2000.
- References
  - C. Manning and H. Schutze, Foundations of Statistical Natural Language Processing, MIT Press, 1999.
  - J. Allen, Natural Language Understanding, Benjamin/Cummings Publishing Co, 1995
  - 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

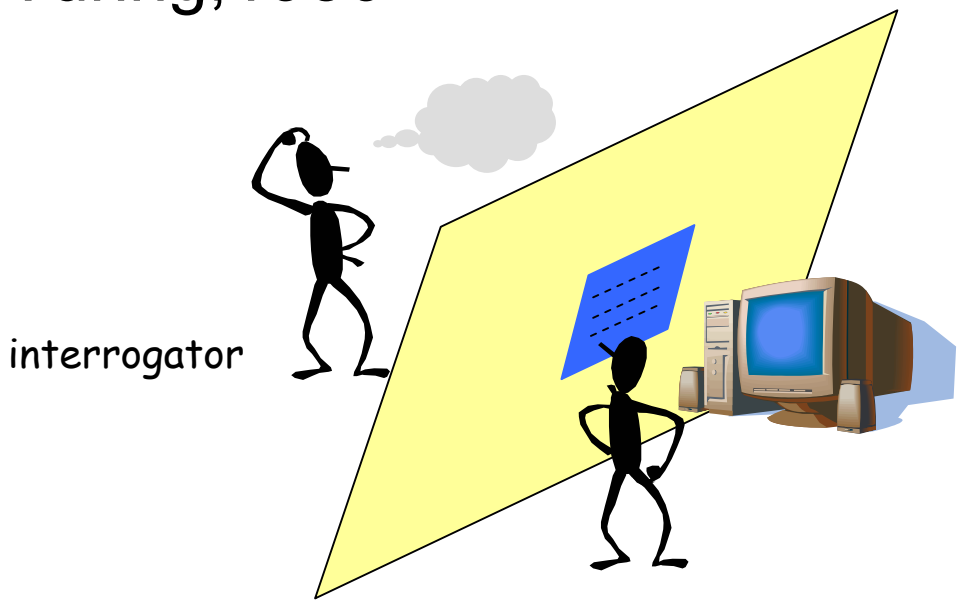
- Different Academic Disciplines: Problems and Methods
  - Electrical Engineering, Statistics
  - Computer Science
  - Linguistics
  - Psychology



- 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

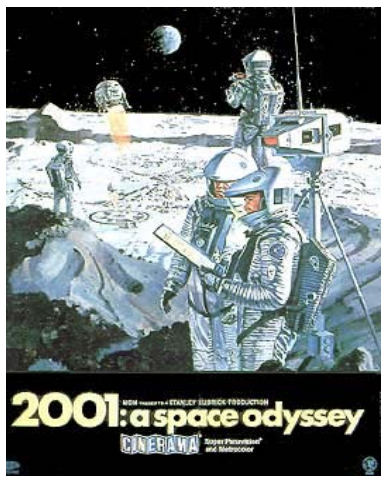
- Alan Turing,1950



  - Predicted at the end of 20 century a machines 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?

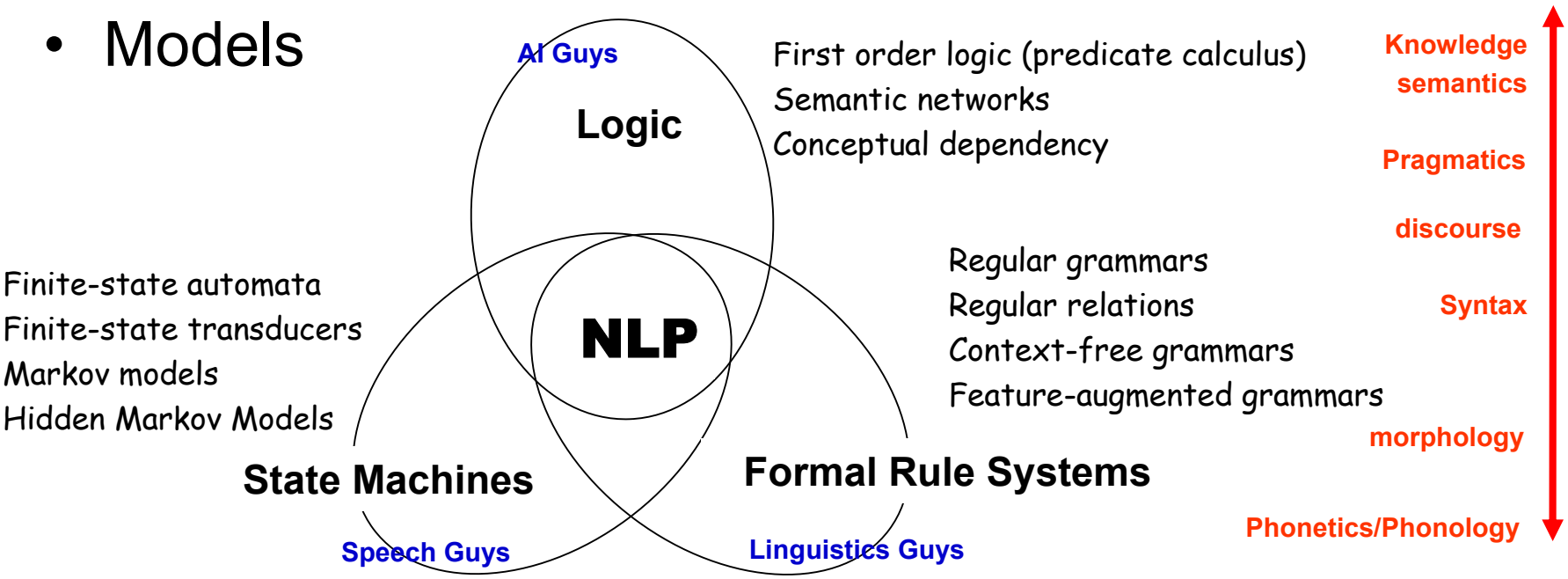
- Read student essays and grade them
- Automated reading tutor
- ……

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
  - 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



- 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
  - Regular expressions
  - Morphology
  - Automata, Finite-State Transducers
- Syntax
  - Part-of-Speech Tagging
  - 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

Applications of NLP

- Speech Recognition
- Information Retrieval and Extraction
- Summarization
- Question Answering
- Conversational Agents
- Machine (Speech/Language) Translation
- Spelling Check
- Segmentation
- 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
  - Taiwan
    - 中研院：陳克健、鄭秋豫、黃居仁、魏培泉 、簡立峰、王新民
    - 台大：陳信希、陳光華；李琳山
    - 成大：王駿發、吳宗憲、簡仁宗
    - 交大：陳信宏、王逸如；張文輝
    - 清大：王小川；張俊盛；張智星
    - 中央：楊接期；張嘉惠
    - 政大：劉昭麟
    - 台大：高照明
    - 台師大：陳柏琳

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**