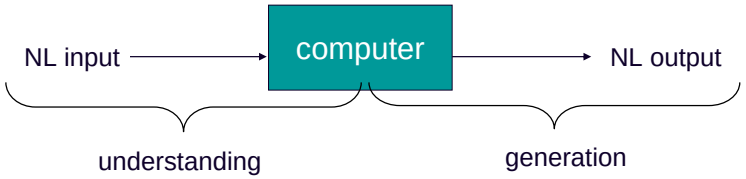
Last class: Why study NLP?



- Useful applications
- Interdisciplinary
- Challenging

Last class: Why is it hard?

Topics for Today

- Brief history of NLP
- Introduction to lexical semantics
- Writing critiques

Early Roots: 1940's and 1950's

- Work on two foundational paradigms
  - Automaton
    - Turing's (1936) model of algorithmic computation
    - Kleene's (1951, 1956) finite automata and regular expressions
    - Shannon (1948) applied probabilistic models of discrete Markov processes to automata for language
    - Chomsky (1956)
    - First considered finite-state machines as a way to characterize a grammar
  - Led to the field of formal language theory

Early Roots: 1940's and 1950's

- Work on two foundational paradigms
  - Probabilistic or information-theoretic models for speech and language processing
    - Shannon: the "noisy channel" model
    - Shannon: borrowing of "entropy" from thermodynamics to measure the information content of a language

Two Camps: 1957-1970

- Symbolic paradigm
  - Chomsky
    - » Formal language theory, generative syntax, parsing
    - » Linguists and computer scientists
    - » Earliest complete parsing systems
      - ◆ Zelig Harris, UPenn
      - ◆ …A possible critique reading!!

Two Camps: 1957-1970

- Symbolic paradigm
  - Artificial intelligence
    - » Created in the summer of 1956
    - » Two-month workshop at Dartmouth
    - » Focus of the field initially was the work on reasoning and logic (Newell and Simon)
    - » Early natural language systems were built
      - ◆ Worked in a single domain
      - ◆ Used pattern matching and keyword search

Two Camps: 1957-1970

- Stochastic paradigm
  - » Took hold in statistics and EE
  - » Late 50's: applied Bayesian methods to OCR
  - » Mosteller and Wallace (1964): applied Bayesian methods to the problem of authorship attribution for *The Federalist* papers.

Additional Developments

- 1960's
  - First serious testable psychological models of human language processing
    - Based on transformational grammar
  - First on-line corpora
    - The Brown corpus of American English
      - 1 million word collection
      - Samples from 500 written texts
      - Different genres (news, novels, non-fiction, academic,….)
      - Assembled at Brown University (1963-64, Kucera and Francis)
    - William Wang's (1967) DOC (Dictionary on Computer)
      - On-line Chinese dialect dictionary

1970-1983

- Explosion of research
  - Stochastic paradigm
    - Developed speech recognition algorithms
      - HMM's
      - Developed independently by Jelinek et al. at IBM and Baker at CMU
  - Logic-based paradigm
    - Prolog, definite-clause grammars (Pereira and Warren, 1980)
    - Functional grammar (Kay, 1979) and LFG

1970-1983

- Explosion of research
  - Natural language understanding
    - SHRDLU (Winograd, 1972)
    - The Yale School
      - Focused on human conceptual knowledge and memory organization
    - Logic-based LUNAR question-answering system (Woods, 1973)
  - Discourse modeling paradigm

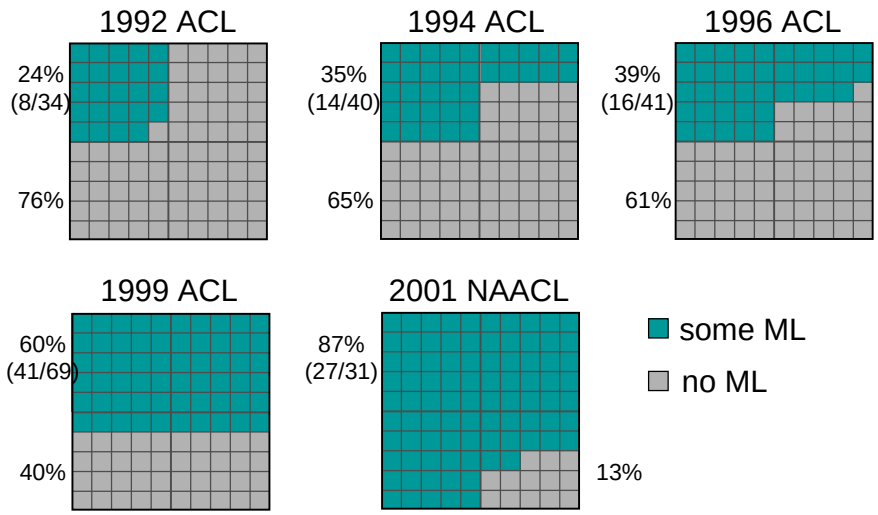
Revival of Empiricism and FSM's

- 1983-1993
  - Finite-state models
    - Phonology and morphology (Kaplan and Kay, 1981)
    - Syntax (Church, 1980)
  - Return of empiricism
    - Rise of probabilistic models in speech and language processing
    - Largely influenced by work in speech recognition at IBM
  - Considerable work on natural language generation

A Reunion of a Sort…

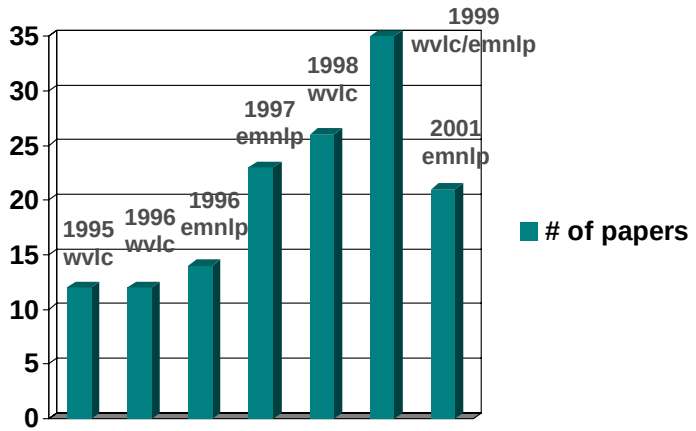
- 1994-pres
  - Probabilistic and data-driven models had become quite standard
  - Increases in speed and memory of computers allowed commercial exploitation of speech and language processing
    - » Spelling and grammar checking
  - Rise of the Web emphasized the need for language-based information retrieval and information extraction

Statistical and Machine Learning Approaches Rule!

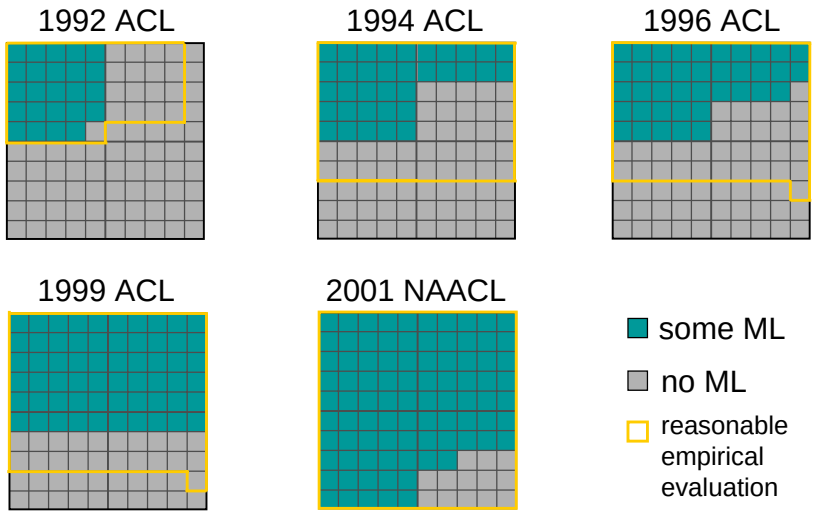


WVLC and EMNLP Conferences

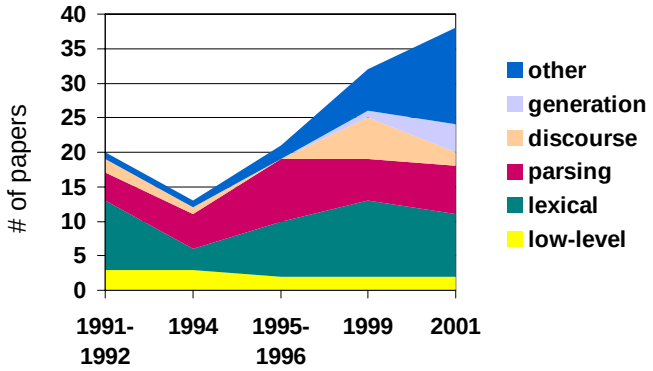
- Workshop on Very Large Corpora
- Conference on Empirical Methods in NLP



Empirical Evaluation



Progression of NL learning tasks



Topics for Today

- Brief history of NLP
- Introduction to lexical semantics
- Writing critiques

Semantic analysis

- Assigning meanings to linguistic utterances
- **Compositional semantics**: we can derive the meaning of the whole sentence from the meanings of the parts.
  - Max ate a green apple.
- Relies on knowing:
  - the meaning of individual words
  - how the meanings of individual words combine to form the meaning of groups of words
  - how it all fits in with syntactic analysis

Caveats

- Problems with a compositional approach
  - a former congressman
  - a toy elephant
  - kicked the bucket

Introduction to lexical semantics

- Lexical semantics is the study of
  - the systematic meaning-related connections among words and
  - the internal meaning-related structure of each word
- Lexeme
  - an individual entry in the lexicon
  - a pairing of a particular orthographic and phonological form with some form of symbolic meaning representation
- Sense: the lexeme's meaning component
- Lexicon: a finite list of lexemes

Dictionary entries

- right *adj.* located nearer the right hand esp. being on the right when facing the same direction as the observer.
- left *adj.* located nearer to this side of the body than the right.
- red *n.* the color of blood or a ruby.
- blood *n.* the red liquid that circulates in the heart, arteries and veins of animals.

Next class

- Providing an NLP system with a large enough knowledge base of such facts will enable it to perform fairly sophisticated semantic tasks (even if the system doesn't know its right from its left).

Topics for Today

- Brief history of NLP
- Introduction to lexical semantics
- ⇨ Writing critiques

Critique Guidelines

- <=1 page, typed (single space)
- The purpose of a critique is **not** to summarize the paper; rather you should choose one or two points about the work that you found interesting.
- Examples of questions that you might address are:
  - What are the strengths and limitations of its approach?
  - Is the evaluation fair? Does it achieve it support the stated goals of the paper?
  - Does the method described seem mature enough to use in real applications? Why or why not? What applications seem particularly amenable to this approach?
  - What good ideas does the problem formulation, the solution, the approach or the research method contain that could be applied elsewhere?
  - What would be good follow-on projects and why?

Critique Guidelines

  - Are the paper's underlying assumptions valid?
  - Did the paper provide a clear enough and detailed enough description of the proposed methods for you to be able to implement them? If not, where is additional clarification or detail needed?

- Avoid **unsupported** value judgments, like ``I liked...'' or ``I disagreed with...'' If you make judgments of this sort, explain why you liked or disagreed with the point you describe.
- Be sure to distinguish comments about the writing of the paper from comment about the technical content of the work.