CMSC 325
Computational Linguistics
Fall 2018
Deepak Kumar
Computational Linguistics

- Study what goes into getting computers to perform useful and interesting tasks involving human languages

- Also concerned with the insights that such computational work gives us into human processing of language
Why care?

• Enormous amount of knowledge is now available in machine readable form as natural language text.

• Conversational agents are becoming common: Siri, Google Voice, Alexa, etc.

• Much of human communication is now mediated by computers.
Topics

• Words
• Syntax
• Meaning
• Discourse
Topics

• Words
• Syntax
• Meaning
• Discourse

Applications exploiting each
Applications – Language Processing versus Data Processing?

• An application that requires the use of **knowledge about human languages**

Example: Is Linux/Unix wc (word count) an example of a language processing application?
Applications – Language Processing versus Data Processing?

• An application that requires the use of **knowledge about human languages**

Example: Is Linux/Unix **wc** (word count) an example of a language processing application?

  • When it counts words: **Yes**
    • To count words you need to know what a word is. That is knowledge of language.

  • When it counts lines and bytes: **No**
    • Lines and bytes are computer artifacts, not linguistic entities.
Some big applications requiring knowledge of language

- Question answering
- Conversation agents
- Summarization
- Machine Translation

These require a tremendous amount of knowledge of language.
Example

• Siri:

  What is the population of Bryn Mawr?
  What should I eat today?
  Tell me a joke.
What knowledge is needed?

• Speech recognition & synthesis
  Knowledge of English words (e.g. what they mean,...)

• How groups of words “clump”
  • What the clumps mean?
Course Content

• Linguistic topics
  • Phonology, morphology, syntax, semantics, discourse structure

• Formal Systems
  • Regular languages, context-free grammars, logic, etc.

• Applications
The Pipeline

- Phonology
- Morphology
- Syntax
- Semantics
- Pragmatics
- Discourse

Sound Waves → Tokens → Words → Parses → Meaning → More Meaning → Discourse → Meaning in context
Ambiguity

• Computational Linguists are obsessed with ambiguity

• It is a fundamental problem of computational linguistics

• Resolving ambiguity is a crucial goal
Linguistic Knowledge, Models & Algorithms

• **Linguistic Knowledge** – characteristics of language (observed/captured phenomena
  Words, morphology, parts of speech, grammars, types of sentences, semantics, etc.

• **Models** – formalisms that are used to capture the various kinds of linguistic knowledge
  that we need.

  Regular Languages, State machines, Rule-based approaches, N-Grams, Logical
  formalisms, Probabilistic models, HMMs, λ-reductions, etc.

• **Algorithms** – used to manipulate the knowledge representations

  Regular expressions, tokenization, Transducers/filters, morphological parsing, state-
  space search, dynamic programming, classifiers, semantic analysis, etc.
The Pipeline

- Phonology
- Morphology
- Syntax
- Semantics
- Pragmatics
- Discourse

Sound Waves → Phonology → Morphology → Syntax → Semantics → Pragmatics → Discourse → Meaning in context
Topics for Exam 3

• CFGs
• Parsing: top-down, bottom up, recursive descent, shift-reduce, CKY, Earley, ATNs.
• Meaning Representations: FOPC
• Semantic Analysis: Syntax-driven, CFGs with semantic attachment $\lambda$-reductions.

Reading: Chapter 13, 17, and 18.