Computation and Linguistic Analysis

L545
Spring 2010
Computation and Linguistic Analysis

From a practical perspective, computational linguistics provides the computational means to deal with spoken and written natural language:

- information extraction
- automatic translation
- natural language frontends to databases
- message generation
- spelling correction
- ...

From a linguistic perspective, computational linguistics provides a possibility to

- formalize and computationally test linguistic theories
- obtain example data relevant to linguistic theorizing

People differ as to what extent they do science vs. engineering ...
Why study CL?

• Human language is interesting and challenging
  – NLP offers insight into properties of language
  – Combines different thinking: ling, CS, psych, math

• Language is the medium of the web

• CL analysis can help in communication

• The field is ambitious yet practical
  – e.g., Machine Translation (MT) is enormously difficult, but MT is useful (e.g., google’s “Translate this page”)
Symbolic CL

This course will focus more on so-called symbolic CL, or theory-driven computational linguistics

- When possible, we will connect the material to theoretical insights
- Although we’ll touch on some statistical methods, those are largely left for L645
General themes

This course will focus on what we need to know in order to represent language

• i.e., a focus on the underlying machinery more than the applications

• If you know how to use FSAs for morphology, you can learn how to use them for information extraction

Some themes that will recur this semester:

• Language is highly ambiguous
• Language processing must therefore be efficient
• Language can be treated as a formal system
Course emphasis

In dealing with language this semester, we will emphasize the following three aspects:

- data structures
- formalisms for expressing grammars using these data structures
- (parsing) algorithms for processing with those grammars
What you need to know (1)

Since we’re focusing on theory, it helps to know a bit about the following:

- Morphology: what are the components of words?
- Syntax: what are the relevant types of constructions in language?
  - Arguments/Adjuncts, Control/raising, UDCs, Anaphora
- Semantics: how do word meanings compositionally form sentential meanings?

We’ll focus a lot on parsing this semester, so some basic knowledge of syntax will help

- If you lack this, let me know asap, and I’ll give you some readings ...
What you need to know (2)

We’re also going to be treating language as a formal system, meaning:

- Language is a set of strings
- A language model recognizes or generates a set of strings

I want you to be comfortable with the idea of representing things formally (mathematically)

- We will deal with set theory
- Again, I can give you some readings, but we will also explicitly discuss it relatively soon.