

Computation and Linguistic Analysis

L445/L545

Dept. of Linguistics, Indiana University
Spring 2017

Computation and Linguistic Analysis

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

- ▶ information extraction & text mining
- ▶ automatic translation
- ▶ message generation
- ▶ dialogue interaction
- ▶ sentiment analysis
- ▶ ...

From a linguistic perspective, computational linguistics provides a possibility to :

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

In other words: Why are you here?

- ▶ Human language is interesting and challenging
 - ▶ NLP offers insight into properties of language
 - ▶ Combines different thinking: ling, CS, psych, math, etc.
- ▶ 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., webpage translations)

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

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 can be modeled as a formal system
- ▶ Much language knowledge can be encoded by hand
- ▶ Language processing must be efficient

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

- ▶ Data structures & models
- ▶ Formalisms for expressing grammars using these data structures
- ▶ Algorithms for processing with those grammars
 - ▶ Syntactic parsing will be of great emphasis

What you need to know (1)

As we focus on theory, it helps to know 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?

What you need to know (2)

We will treat 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**