

# INTRODUCTION TO NATURAL LANGUAGE PROCESSING

## INTRODUCTION

# Natural Language Processing

- The field of *Natural Language Processing (NLP)*, or *Computational Linguistics (CL)*, or *Human Language Technology (HLT)*, is primarily concerned with the creation of computer programs that perform useful and interesting tasks with human languages (e.g., understanding, generation, learning).
- It is secondarily concerned with using computational metaphors to help us come to a better understanding of human language.
- The foundations of the field are in computer science (e.g., AI, theory), linguistics, mathematics and statistics, electrical engineering, and psychology.
- Studying NLP involves studying natural languages, formal representations, and algorithms for their manipulation.

# Major Topics of this Course

## Knowledge of Language

- words (meaningful components of words)
- syntax (structure of a sentence)
- semantics (explicit meaning of a sentence)
- pragmatics/discourse (implicit/contextual meaning)

## Methodologies and Tools

- knowledge-based and statistical
- state machines, rule systems, grammars, logic, search, probability, automata, machine learning, and more

## Applications

- from hyphenators to intelligent agents

## Current Research

# Knowledge of Language

- Example dialogue from *2001: A Space Odyssey*

Dave: Open the pod bay doors, HAL.

HAL: I'm sorry Dave, I'm afraid I can't do that

- To participate in such a conversation, HAL needs knowledge about many levels of language
  - words: producing contractions, plurals
  - syntax: questions versus statements, word order and grouping
  - semantics: meaning of words in isolation and compositionally
  - pragmatics: politeness and indirectness
  - discourse: between utterance references

# Applications

What makes a computer application a language processing application?

- language processing applications require the use of knowledge of language

# Little Applications

Little applications typically make use of only a small amount of a single kind of knowledge of language. Many of these types of applications are currently in use, but they are often nearly invisible.

- line breakers
- hyphenators
- spelling correctors
- OCR software
- grammar and style checkers

# Big Applications

Big applications often make use of large amounts and varied kinds of knowledge of language.

- information retrieval and extraction
- question answering
- dialogue management
- text summarization
- machine translation
- HAL