

UNIVERSITY of DELAWARE

UNIVERSITY of DE

Basics of Natural Language Processing

(Search and Data Mining)

Hui Fang
Department of Electrical and Computer Engineering
University of Delaware

1

UNIVERSITY of DELAWARE

UNIVERSITY of DE

What is Natural Language Processing (NLP)?

你会讲中文吗?

How can a computer make sense out of this string?

- Morphology** - What are the basic units of meaning (words)?
- Syntax** - How are words related with each other?
- Semantics** - What is the "combined meaning" of words?
- Pragmatics** - What is the "meta-meaning"? (speech act)
- Discourse** - Handling a large chunk of text
- Inference** - Making sense of everything

UNIVERSITY of DELAWARE

UNIVERSITY of DE

An Example of Tasks in NLP

A dog is chasing a boy on the playground

Det Noun Aux Verb Det Noun Prep Det Noun

Noun Phrase Complex Verb Noun Phrase Prep Phrase

Semantic analysis

Dog(d1). Boy(b1). Playground(p1). Chasing(d1,b1,p1).

Sentence

Syntactic analysis (Parsing)

Pragmatic analysis (speech act)

A person saying this may be reminding another person to get the dog back...

Inference

Scared(b1)

UNIVERSITY of DELAWARE

UNIVERSITY of DE

NLP is Difficult!!!!!!

- Natural language is designed to make human communication efficient.
 - We omit a lot of "common sense" knowledge, which we assume the hearer/reader possesses
 - we keep a lot of ambiguities, which we assume the hearer/reader knows how to resolve
- This makes EVERY step in NLP hard
 - Ambiguity is a "killer"!
 - Common sense reasoning is pre-required

UNIVERSITY of DELAWARE

UNIVERSITY of DE

Examples of Challenges

- Word-level ambiguity:
 - "design" can be a noun or a verb (Ambiguous POS)
 - "root" has multiple meanings even as a noun (Ambiguous sense)
- Syntactic ambiguity:
 - "natural language processing" (Ambiguous modification)
 - "A man saw a boy with a telescope." (Ambiguous PP Attachment)
- Anaphora resolution:
 - "John persuaded Bill to buy a TV for himself." (himself = John or Bill?)
- Presupposition:
 - "He has quit smoking." implies that he smoked before.

UNIVERSITY of DELAWARE

UNIVERSITY of DE

Despite all the challenges, research in NLP has also made a lot of progress...

