Structure of Language

Natural Language Processing
Language Technology
Week 1

Andrew Rosenberg
CUNY Graduate Center
Computational Linguistics
or
Natural Language Processing
What do we do in NLP?

Human Machine Communication
What do we do in NLP?

Human-Human Communication
Web Search

- Information Retrieval
Question Answering
Language Generation

Scott Horsley, Wicked Fast Reporter

Machine Translation

I was wondering what you are going to do later?
Me preguntaba lo que vas a hacer después?

Going to the pub, do you want to join us? I think you met some of the team at the party in April.
Al pub, ¿quisieras unirte a nosotros? Creo que conociste a algunos miembros del equipo en la fiesta en abril.

Can I join you guys after my meeting?
Puedo unirme a ustedes después de mi reunión?
Speech Recognition
Conversational Agents

- Chat Bots
- Customer Service
- Virtual Assistants
- Tutoring Systems
What else?
Structure in Language

- Pragmatics
- Semantics
- Syntax
- Words
- Morphology
- Phonology
- Phonetics
- Paralinguistics
- Letters
Syntax

- Formal relationships between words.
- Defines equivalence classes i.e. “parts of speech”.
- Modification and attachment rules
- Phrase structure
Semantics

- The study of “meaning”
- Usually “literal meaning”
- Formal, abstract representations
Pragmatics

• Meaning that involves:
  • Speaker
  • Hearer
  • Context
Morphology

• Study of meaningful components of words

• Sub-word units

• Inflectional
  • usually responsible for filling a syntactic function e.g. pluralizing or agreement

• Derivational
  • creates a new related word, e.g. “computer” “computer -ize” “computerize -ation”
Letters and Spelling

• Languages have spelling rules and/or conventions

• Stems and morphemes have systematized spellings.

denationalization
Phonology

- Categorical inventory of speech sounds.

- Pronunciation of words/morphemes/letters.

  letter to sound (grapheme to phoneme)

[ai pʰiː eɪ]
Phonetics

• The relationship between acoustics and language (phonology)

• Largely not covered in this class. Requires some (digital) signal processing that we won’t have time for
Paralinguistics

- Information that’s not related to literal meaning that is conveyed via language.
- Speaker state: Emotions, Fatigue, Intoxication
- Personal Attributes: Age, Gender
What do we mean when we say “language”
• James Pennebaker: Pronouns reveal a surprising amount particularly w.r.t. **pragmatics** and **paralinguistics**.

• Shallow Analysis
  • Analysis of words, characters, etc.

• Deep Analysis
  • Syntactic and semantic structure
  • Incorporation of external knowledge
What makes language processing difficult?
Ambiguity

• I live at the bank
  • Lexical

• I saw the man with the telescope
  • Structural — “attachment ambiguity”

• I saw her duck
  • Lexical / Semantic / Syntactic

• A: Do you like Chinese food?
  B: I like Italian.
  • Pragmatic/discourse — i.e. Why are you telling me this? What are you trying to say here.
Relationship to human language

Output vs. Process
Parts of Speech

Is bank a noun or a verb?

• How would you decide?
Machine Learning

- Machine Learning gives a mechanism for relationships between observations and a class label.

  - e.g. Words and parts of speech

\[ f(x) = t \]

- Claim:

  - An algorithm with a lot of examples to learn from can train a better decision function than a human.
Classification
Structured Prediction

\[ f(x) = t \]

• What if \( t \) isn’t a just an element of a set, but a more complicated object, with relationships to other objects (i.e., structure)?

• \( x = \) a sequence of words
  \( t = \) a parse tree

• \( x = \) a sequence of words
  \( t = \) another sequence of words

How does this change the process?
Homework #1, Part 1

• **Reading**
  
  • Read J&M Chapter 2.
  
  • Bring a question or comment to class — written.

• **Task**

  • Find some dialog involving a robot.
  
  • Describe in as much detail as you can, what capabilities the robot needs to have to 1) understand its conversational partner and 2) generate its responses.