Natural Language Processing

Gholamreza Ghassem-Sani
Main References:

1. Natural Language Understanding, James Allen, 2nd ed.
2. Speech and Language Processing, Daniel Jurafsky and James Martin, 2nd ed.
What is Natural Language Processing?

computers using natural language as input and/or output

language → computer → language

understanding (NLU) → generation (NLG)
Goals of the field

Develop computational models of human language processing

• ...in order to write computer programs performing smart tasks with language
• ...in order to gain a better understanding of the human language processor, and of communication
Interdisciplinary research...

• Linguistics
  – How do words form phrases and sentences?

• Psychology
  – How do people identify the meaning of words and structure of sentences?

• Philosophy
  – What is the meaning, and how do words and sentences acquire it?

• Artificial Intelligence
  – How is the structure of sentences identified? How can knowledge and reasoning be modeled?
NLP: applications

- Speech recognition and synthesis
- Machine translation
- Document processing
  - information extraction
  - summarization
- Text generation
- Dialog systems (typed and spoken)
**Speech recognition**

- Task: From representation of audio stream to sequence of words
- Word hypotheses lattice
- Applications: dictation, ...
Machine Translation

SL-Text

Analysis

Transfer

Interlingua

Generation

TL-Text
Translating Speech

- Translate negotiation dialogs between German, English, Japanese in near-real-time
- Process spontaneous speech
On Monday, GreenChip Solutions made an acquisition offer to BuyOut Inc., a St. Louis-based plastic tree manufacturer that had tremendous success in equipping American households with pink plastic oak trees.

Analysis

Generation

GreenChip offered to acquire the plastic tree manufacturer BuyOut.
Information Extraction

• Dedicated search for key information in the text
  – acquirer: GreenChip
  – acquiree: BuyOut
  – date-offer: Monday
  – $$-offer: UNSPEC

• Key attributes are predefined

• Systems are domain-specific
Text generation

• Reports
  – Data mining results
  – Project status
  – Continuous data (e.g., air pollution)

• Technical Documentation
  – User manuals
  – multilingual, many variants, frequent updates
Dialog systems

• Web-based, typed
• Spoken: telephone (web)
  – Travel timetable
  – Cinema programs
• E-Commerce: product search & information
Levels of language analysis

- **Phonetics/phonology**: what words (or sub words) are we dealing with?
- **Morphological knowledge**: how words are constructed from more basic meaning units called *morphemes*?
- **Syntax**: What phrases are we dealing with? Which words modify one another?
- **Semantics**: What’s the literal meaning (i.e., context-free meaning)?
- **Pragmatics**: What should you conclude from the fact that I said something?
- **Discourse Knowledge**: how the immediately preceding sentences affect the interpretation of the next sentence?
- **World knowledge**: includes the general knowledge about the world (i.e., knowledge about others’ beliefs and goals)
Levels of language analysis

- Phonetics: sounds -> words
  - /b/ + /o/ + /t/ = boat
- Morphology: morphemes -> words
  - friend + ly = friendly
- Syntax: word sequence -> sentence structure
- Semantics: sentence structure + word meaning -> sentence meaning
- Pragmatics: sentence meaning + context -> more precise meaning
- Discourse and world knowledge
Levels of language analysis (cont.)

1. Language is one of fundamental aspects of human behavior and is crucial component of our lives.
2. Green frogs have large noses.
3. Green ideas have large noses.
4. Large have green ideas nose.
5. I go store.
“Symbolics or statistics”

- Statistic used for:
  - speech recognition
  - part-of-speech tagging
  - parsing
  - machine translation, info retrieval, summarization

- Symbolic reasoning used for:
  - parsing
  - semantic analysis
  - generation
  - machine translation

- Hybrid approaches
Parsing

Utterance

Grammar

Syntactic analysis

Semantic analysis

Pragmatic analysis

Lexicon
Syntactic Analysis

• By analyzing sentence structure (parsing), determine how words are related to each other
  – (The big furry rabbit) (jumped over (the lazy tortoise))

• Obtain the parse tree to represent groupings
Syntactic analysis (2)

• Syntax can make explicit when there are several possible interpretations
  – *(Rice flies) like sand.*
  – *Rice (flies like sand).*

• Knowledge of ‘correct’ grammar can help finding the right interpretation
  – *Flying planes are dangerous.*
  – *Flying planes is dangerous.*
Semantic Analysis

• Different sentences with same meaning should have the same representation:
  – *John gave the book to Mary.*
  – *The book was given to Mary by John.*
  – give-action:
    agent: John
    object: book
    receiver: Mary

• Then we can do reasoning, answer questions
The Flow of Information

---

Figure 1.5 The flow of information
Why is NLP Hard?

“At last, a computer that understands you like your mother”
Ambiguity

- “At last, a computer that understands you like your mother”
  1. (* ) It understands you as well as your mother understands you
  2. It understands (that) you like your mother
  3. It understands you as well as it understands your mother
- 1 and 3: Does this mean well, or poorly?
Ambiguity at Many Levels

At the acoustic level (speech recognition):

1. “... a computer that understands you like your mother”
2. “... a computer that understands you lie cured mother”
Ambiguity at Many Levels

At the syntactic level:

Different structures lead to different interpretations.
More Syntactic Ambiguity

VP
  V
  list

VP
  NP
    DET
    all

VP
  NP
    PP

flights
on Tuesday
At the semantic (meaning) level:
Two definitions of “mother”
• a woman who has given birth to a child
• a stringy slimy substance consisting of yeast cells and bacteria; is added to cider to produce vinegar (i.e., mother of vinegar)
This is an instance of word sense ambiguity
More Word Sense Ambiguity

At the semantic (meaning) level:

• They put money in the bank  
  = buried in mud?

• I saw her duck with a telescope
Ambiguity at Many Levels

At the discourse level:

- Alice says they’ve built a computer that understands you like your mother
- But she ...
- ... doesn’t know any details
- ... doesn’t understand me at all
John stopped at the donut store on his way home from work. He thought a coffee was good every few hours. But it turned out to be too expensive there.
John stopped at the *donut store* on his way home from work. He thought a coffee was good every few hours. But it turned out to be too expensive there.

To get a *donut* (spare tire) for his car?
What’s hard about this story?

John stopped at the *donut store* on his way home from work. He thought a coffee was good every few hours. But it turned out to be too expensive there.

store where donuts shop? or is run by donuts? or looks like a big donut? or made of donut? or has an emptiness at its core?
What’s hard about this story?

I stopped smoking freshman year, but John stopped at the donut store on his way home from work. He thought a coffee was good every few hours. But it turned out to be too expensive there.
What’s hard about this story?

John stopped at the donut store on his way home from work. He thought a coffee was good every few hours. But it turned out to be too expensive there.

Describes where the store is? Or when he stopped?
What’s hard about this story?

John stopped at the donut store on his way home from work. He thought a coffee was good every few hours. But it turned out to be too expensive there.

Well, actually, he stopped there from hunger and exhaustion, not just from work.
What’s hard about this story?

John stopped at the donut store on his way home from work. He thought a coffee was good every few hours. But it turned out to be too expensive there.

At that moment, or habitually?

(Similarly: Mozart composed music.)
John stopped at the donut store on his way home from work. He thought a coffee was good every few hours. But it turned out to be too expensive there.

That’s how often he thought it?
What’s hard about this story?

John stopped at the donut store on his way home from work. He thought a coffee was good every few hours. But it turned out to be too expensive there.

But actually, a coffee only stays good for about 10 minutes before it gets cold.
What’s hard about this story?

John stopped at the donut store on his way home from work. He thought a coffee was good every few hours. But it turned out to be too expensive there.

Similarly: In America a woman has a baby every 15 minutes. Our job is to find that woman and stop her.
What’s hard about this story?

John stopped at the donut store on his way home from work. He thought a coffee was good every few hours. But it turned out to be too expensive there.

the particular coffee that was good every few hours? the donut store? the situation?
What’s hard about this story?

John stopped at the donut store on his way home from work. He thought a coffee was good every few hours. But it turned out to be too expensive there.

too expensive for what? what are we supposed to conclude about what John did?
Turing test

AI system

Control

Experimenter
Patient: You are like my father in some ways.
Doctor: What resemblance do you see?
Patient: You are not very aggressive.
Doctor: What makes you think I am not very aggressive?
Patient: You don’t argue with me.
Doctor: Why do you think I don’t argue with you?
Patient: You are afraid of me.
Doctor: Does it please you to believe I am afraid of you?
Patient: My father is afraid of everybody.
Doctor: What else comes to mind when you think of your father?
Patient: Bullies.
The Chinese Room (Searle)

Set of rules, in English, for transforming phrases

Chinese Writing is given to the person

She does not know Chinese

Correct Responses