Overview: Natural Language Processing

What is Natural Language Processing?

- The study of human languages and how they can be represented computationally and analyzed and generated algorithmically
  - The cat is on the mat → on (mat, cat)
  - on [mat, cat] → The cat is on the mat

- Studying NLP involves studying natural language, formal representations, and algorithms for their manipulation

What is Natural Language Processing?

- Building computational models of natural language for comprehension and production

Other Names:
- Computational Linguistics (CL)
- Human Language Technology (HLT)
- Natural Language Engineering (NLE)
- Speech and Text Processing
What is Natural Language Processing?

Engineering Perspective

- Use NLP as part of a larger application:
  - Spoken dialogue systems for telephone services
  - Components of web search or document retrieval services
    - Machine translation
    - Question/answering systems
    - Text summarization
  - Interface for intelligent tutoring/training systems
- Emphasis on:
  - Robustness (doesn’t collapse on unexpected input)
  - Coverage (does something useful with most inputs)
  - Efficiency (speech, large document collections)

Cognitive Science Perspective

- Goals
  - Gain an understanding of how people comprehend and produce language.
  - A model that explains actual human behaviour.

Solution must:
- Explain psycholinguistic data
- Be verified by experimentation
What is Natural Language Processing?

Theoretical Linguistics Perspective

- In principle, coincides with the Cognitive Science Perspective
  - CL can potentially help test the empirical adequacy of theoretical models.
- Linguistics is typically a descriptive enterprise
  - Building computational models of the theories allows them to be empirically tested.
    - E.g., does your grammar correctly parse all the grammatical examples in a given test suite, while rejecting all the ungrammatical examples?

Language as Goal-Oriented Behaviour

- We speak for a reason, e.g.,
  - get hearer to believe something
  - get hearer to perform some action
  - impress hearer

Examples

1. It’s hot in here, isn’t it?
2. Can you book me a flight to London tomorrow morning?
3. P: What time does the train for Washington, DC leave?
   C: 6:00 from Track 17.
Typical Focus

- **Language Production**
  - Must determine how to use linguistic strategies to achieve desired effects
- **Language Understanding**
  - Must use linguistic knowledge to recognise speaker’s underlying purpose

Knowledge needed to produce and understand language

- **Phonetics and phonology**: how words are related to sounds that realize them
- **Morphology**: how words are constructed from more basic meaning units
- **Syntax**: how words can be put together to form correct utterances
- **Lexical semantics**: what words mean
- **Compositional semantics**: how word meanings combine to form larger meanings
- **Pragmatics**: how situation affects interpretation of utterance
- **Discourse structure**: how preceding utterances affects processing of next utterance

Knowledge needed:

**Phonetics and Phonology**

- Speech sounds, their production, and the rule systems that govern their use
  - Sín vs. Sín (him vs. hymn) (when heard)
  - Dagur, Dogg
  - Víll (mannsnafn), Víll (eitthvað rangt)
  - Á bílastæðinu við húsið – Á bílastæðinu við húsið
  - Maturinn kominn! – Maturinn kominn?
Knowledge needed:
**Morphology**

• How words are constructed from more basic units, called morphemes

\[
\text{friend} + \text{ly} = \text{friendly}
\]

noun

Suffix -ly turns noun into an adjective (and verb into an adverb)

Knowledge needed:
**Syntax**

• How words can be put together to form legal sentences in the language
• What structural role each word plays in the sentence
• What phrases are subparts of other phrases

prepositional phrase

The **white book** by Jurafsky and Martin is fascinating.

modifier

noun phrase

modifier

Knowledge needed:
**Semantics**

• What words mean (Lexical Semantics)
• How word meanings combine in sentences to form sentence meanings (Composition Semantics)
• The sole died

Syntax and semantics work together!

1. What does it taste like?
2. What taste does it like?

**Semantics:** Meaning independent of a larger context!
Knowledge needed:

Pragmatic Knowledge

• The meaning of words and phrases in context
  – George got married and had a baby.
  – George got bread and butter.

• Meaning often indicated by intonation/prosody.
  • German teachers
  • Bill doesn't drink because he's unhappy.
  • John only introduced Mary to Sue.
  • John called Bill a Republican and then he insulted him

Knowledge needed:

Pragmatic Knowledge

• What utterances mean in different contexts

Jon was hot and desperate for a dunk in the river.
Jon suddenly realised he didn't have any cash.
He rushed to the bank.

financial institution    river bank

Knowledge needed:

Discourse Structure

• Much meaning comes from simple conventions that we generally follow in discourse
• For example how we refer to entities
  - Indefinite NPs used to introduce new items into the discourse
    A woman walked into the cafe.
  - Definite NPs can be used to refer to subsequent references
    The woman sat by the window.
  - Pronouns used to refer to items already known in discourse
    She ordered a cappuccino.
Knowledge needed:
**Discourse Relations**

- Relationships we infer between discourse entities
- Not expressed in either of the propositions, but from their juxtaposition

(a) Ëg er glorsofrin
(b) Fórum á Sólon

(E.g. (b) is a “solution” in Rhetorical Structure Theory)

---

Knowledge needed:
**Discourse and Temporal Interpretation**

Max fell. John pushed him.

Explanation

**Syntax and semantics:** “him” refers to Max

**Lexical semantics and discourse:** the pushing occurred before the falling.

---

Knowledge needed:
**Discourse and World Knowledge**

- What we know about the world and what we can assume our hearer knows about the world is intimately tied to our ability to use language

_I took the cake from the plate and ate it._
Ambiguity

I made her duck.

• The categories of knowledge of language can be thought of as ambiguity-resolving components
• How many different interpretations does the above sentence have?
• How can each ambiguous piece be resolved?
• Does speech input make the sentence even more ambiguous?

Basic Process of NLU

It’s not that simple

• Syntax affects meaning
  1. (a) Flying planes is dangerous.
     (b) Flying planes are dangerous.

• Meaning and world knowledge affects syntax
  2. (a) Flying insects is dangerous.
     (b) Flying insects are dangerous.

  3. (a) I saw the Grand Canyon flying to LA.
     (b) I saw a condor flying to LA.