Module 13

Natural Language Processing

Version 2 CSE IIT, Kharagpur

Lesson 41

Parsing

Version 2 CSE IIT, Kharagpur

13.3 Natural Language Generation

The steps in natural language generation are as follows.

Meaning representation

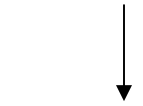
Utterance Planning

Meaning representations for sentences

Sentence Planning and Lexical Choice

Syntactic structures of sentences with lexical choices

Sentence Generation



Morphologically analyzed words

Morphological Generation

Words

13.4 Steps in Language Understanding and Generation

13.4.1 Morphological Analysis

- Analyzing words into their linguistic components (morphemes).
- Morphemes are the smallest meaningful units of language.

cars	car+PLU	
giving	give+PROG	
geliyordum	gel+PROG+PAST+1SG	- I was coming

- Ambiguity: More than one alternatives

flies	flyVERB+PROG	
	flyNOUN+PLU	
adam	adam+ACC	- the man (accusative)
	adam+P1SG	- my man
	ada+P1SG+ACC	- my island (accusative)

13.4.2 Parts-of-Speech (POS) Tagging

- Each word has a part-of-speech tag to describe its category.

- Part-of-speech tag of a word is one of major word groups (or its subgroups).
  - **open classes** -- noun, verb, adjective, adverb
  - **closed classes** -- prepositions, determiners, conjuctions, pronouns, particples

- POS Taggers try to find POS tags for the words.

- duck is a verb or noun? (morphological analyzer cannot make decision).

- A POS tagger may make that decision by looking the surrounding words.
  - Duck! (verb)
  - Duck is delicious for dinner. (noun)

13.4.3 Lexical Processing

- The purpose of lexical processing is to determine meanings of individual words.

- Basic methods is to lookup in a database of meanings – **lexicon**

- We should also identify non-words such as punctuation marks.

- Word-level ambiguity -- words may have several meanings, and the correct one cannot be chosen based solely on the word itself.
  - bank in English

- Solution -- resolve the ambiguity on the spot by POS tagging (if possible) or pass-on the ambiguity to the other levels.

13.4.4 Syntactic Processing

- **Parsing** -- converting a flat input sentence into a hierarchical structure that corresponds to the units of meaning in the sentence.

- There are different parsing formalisms and algorithms.

- Most formalisms have two main components:
  - **grammar** -- a declarative representation describing the syntactic structure of sentences in the language.
  - **parser** -- an algorithm that analyzes the input and outputs its structural representation (its parse) consistent with the grammar specification.

- CFGs are in the center of many of the parsing mechanisms. But they are complemented by some additional features that make the formalism more suitable to handle natural languages.

13.4.5 Semantic Analysis

- Assigning meanings to the structures created by syntactic analysis.
- Mapping words and structures to particular domain objects in way consistent with our knowledge of the world.
- Semantic can play an import role in selecting among competing syntactic analyses and discarding illogical analyses.
  - I robbed the bank -- bank is a river bank or a financial institution
- We have to decide the formalisms which will be used in the meaning representation.

13.5 Knowledge Representation for NLP

- Which knowledge representation will be used depends on the application -- Machine Translation, Database Query System.
- Requires the choice of representational framework, as well as the specific meaning vocabulary (what are concepts and relationship between these concepts -- ontology)
- Must be computationally effective.
- Common representational formalisms:
  - first order predicate logic
  - conceptual dependency graphs
  - semantic networks
  - Frame-based representations

13.6 Discourse

- Discourses are collection of coherent sentences (not arbitrary set of sentences)
- Discourses have also hierarchical structures (similar to sentences)
- **anaphora resolution** -- to resolve referring expression
  - Mary bought a book for Kelly. **She** didn't like **it**.
    - **She** refers to Mary or Kelly. -- possibly Kelly
    - **It** refers to what -- book.
  - Mary had to lie for Kelly. **She** didn't like **it**.

- Discourse structure may depend on application.
  - Monologue
  - Dialogue
  - Human-Computer Interaction

13.7 Applications of Natural Language Processing

- Machine Translation – Translation between two natural languages.
  - See the Babel Fish translations system on Alta Vista.

- Information Retrieval – Web search (uni-lingual or multi-lingual).

- Query Answering/Dialogue – Natural language interface with a database system, or a dialogue system.

- Report Generation – Generation of reports such as weather reports.

- Some Small Applications –
  - Grammar Checking, Spell Checking, Spell Corrector

13.8 Machine Translation

- Machine Translation refers to converting a text in language A into the corresponding text in language B (or speech).

- Different Machine Translation architectures are:
  - interlingua based systems
  - transfer based systems

- Challenges are to acquire the required knowledge resources such as mapping rules and bi-lingual dictionary? By hand or acquire them automatically from corpora.

- Example Based Machine Translation acquires the required knowledge (some of it or all of it) from corpora.

Questions

1. Consider the following short story:

John went to the diner to eat lunch. He ordered a hamburger. But John wasn't very hungry so he didn't _nish it. John told the waiter that he wanted a doggy bag. John gave the waiter a tip. John then went to the hardware store and home.

Each inference below is based on a plausible interpretation of the story. For each inference, briefly explain whether that inference was primarily based on syntactic, semantic, pragmatic, discourse, or world knowledge. (Do not answer world knowledge unless none of the other categories are appropriate.)

(a) John is the person who ordered a hamburger.

(b) John wasn't just stating a fact that he desired a doggy bag, but was requesting that the waiter bring him a doggy bag.

(c) John went to the hardware store and then went to his house. (As opposed to going to a hardware store and a hardware home.)

(d) John gave the waiter some money as a gratuity. (As opposed to giving him a suggestion or hint.)

(e) John was wearing clothes.

2. Identify the thematic role associated with each noun phrase in the sentence below:

Mary went from Utah to Colorado with John by bicycle.

Solutions

1.a. Discourse knowledge. The inference comes from coreference resolution between John" and "He" in the first and second sentences.

1.b. Pragmatics. Most people would assume that John was making a request of the waiter and not merely stating a fact, which is a pragmatic issue because it reects the purpose of John's statement.

1.c. Syntactic knowledge. This inference reflects one syntactic parse: ((hardware store) and (home)), as opposed to an alternative parse: (hardware (store and home)).

1.d Semantic knowledge. Most people would assume that "tip" means gratuity, as opposed to other meanings of the word "tip", such as suggestion or hint.

1.e. World Knowledge. There is nothing stated in the story that mentions clothes, but in our culture people virtually always wear clothes when they leave their house. So we make this assumption.

2. The roles are

agent = Mary
source (from-loc) = Utah
destination (to-loc) = Colorado
co-agent = John
instrument = bicycle