Stages of NLP

Stages of language processing

- Phonetics and phonology
- Morphology
- Lexical Analysis
- Syntactic Analysis
- Semantic Analysis
- Pragmatics
- Discourse

Phonetics

- Processing of speech
- Challenges
  - Homophones: *bank (finance)* vs. *bank (river bank)*
  - Near Homophones: *maatraa* vs. *maatra (hin)*
  - Word Boundary
    - *aajaayenge (aa jaayenge (will come)* or *aaj aayenge (will come today)*
    - *I got [ua]plate*
  - Phrase boundary
    - *mtech1 students are especially exhorted to attend as such seminars are integral to one's post-graduate education*
  - Disfluency: *ah, um, ahem etc.*

1. Phonetics and Phonology

- It is concerned with processing of speech and sound
- concerns processing accents, pauses, amplitude tone etc.

Challenges

Homophones: words which sound similar
e.g bank(finances) v/s bank(river)
bat v/s bat

New Homophones: words which have very close sound

Near Homophones: words which have very close sound

e.g. peas, peace

to, two, too

macrocad, macaca

write, right

knows, nose

word-boundry : break word at proper place

aajaayenge → aaj aayenge
(will come today)

aa jaayenge
(will come)

I got [ua] plate → I got up late

I got a plate

Disfluency: ah, um, ahem etc.

- NO meaning at all, speaker only use to organize her/his thoughts

Morphology

- Word formation rules from *root* words
- Nouns: Plural (*boy-boys);* Gender marking (czar-czarina)
- Verbs: Tense (*stretch-stretched);* Aspect (*e.g. perfective sit-had sat*); Modality (e.g. *request khaanaa→ khaaiie)*
- First crucial first step in NLP
- Languages rich in morphology: e.g., Dravidian, Hungarian, Turkish
- Languages poor in morphology: Chinese, English
- Languages with rich morphology have the advantage of easier processing at higher stages of processing
- A task of interest to computer science: *Finite State Machines for Word Morphology*

Detect all morphenes contained in large word string. Break the words and obtained properties

morphenes : A smallest meaningful unit in the grammar of a language

townhall : "town", "hall"
dogs : "dog", "s"
happiness : "happy", "ness"

Language with rich morphology have the advantage of easier processing at higher stages of processing

Lexical Analysis

- Essentially refers to dictionary access and obtaining the properties of the word

  e.g. dog

  noun (lexical property)
  take-'s'-in-plural (morph property)
  animate (semantic property)
  4-legged (-do-)
  carnivore (-do)

Challenge: *Lexical or word sense disambiguation*

Why would we require such properties?

During morphological stage dogs is divided into 'dog' & 's'. Root word with its properties is stored in dictionary.

To- Embed richness in the entire Data Structure

Dogs-> Morphological Analysis -> Dog-> What is the property of the word Dog-> Animal->4 legged

-For QA based systems, Rich dictionary

Lexical Ambiguity

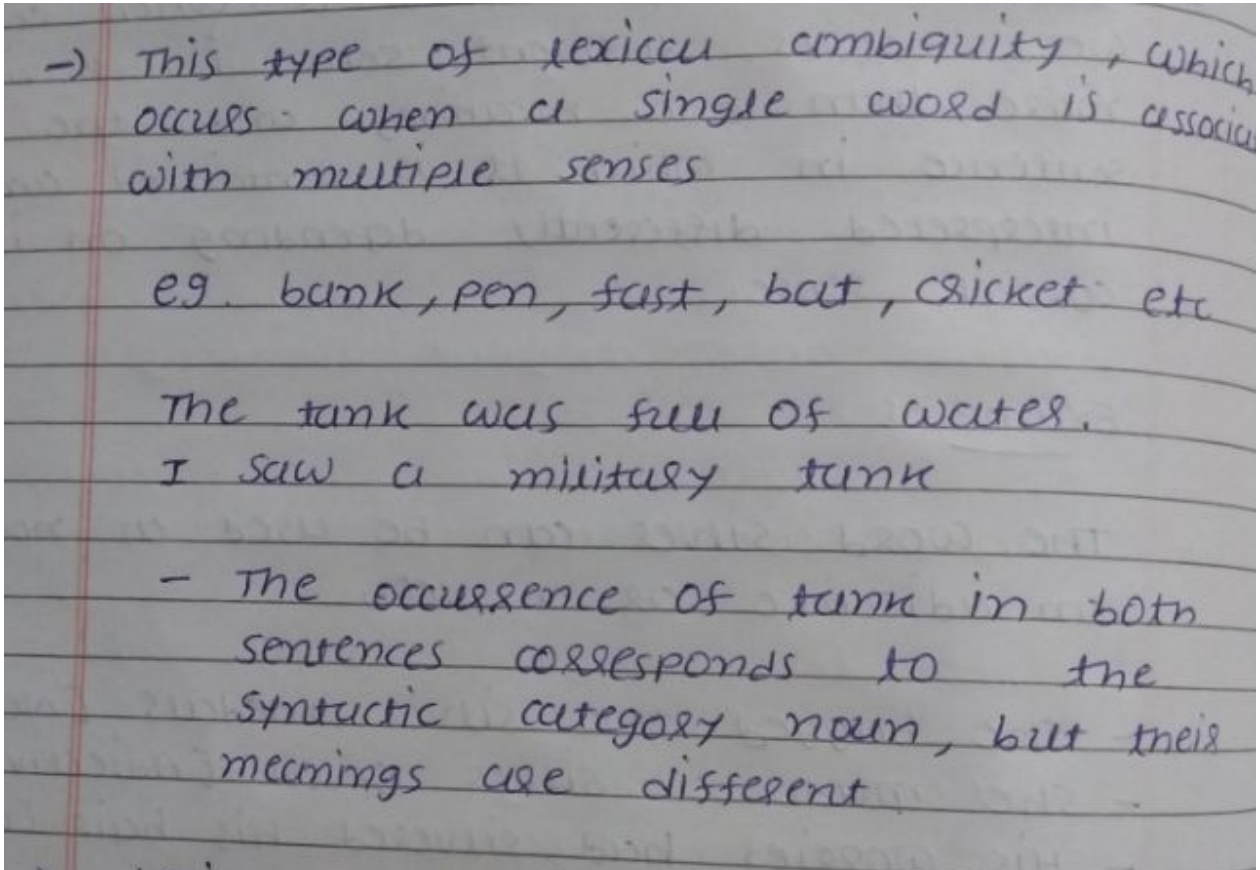
- Lexical ambiguity can occur when a word carries different sense i.e. having more than one meaning and the sentence in which it is contained can be interpreted differently depending on its correct sense

Examples

The word silver can be used as noun, an adjective or a verb

- She bagged two silver medals [Noun]
- She made a silver speech [Adjective]
- His worries had silvered his hair [verb]

Lexical Semantic Ambiguity



Word Sense Disambiguation

Lexical Disambiguation

First step: *part of Speech Disambiguation*

- *Dog* as a *noun* (animal)
- *Dog* as a verb (*to pursue)*

Sense Disambiguation

- *Dog (*as *animal)*
- *Dog (*as *a very detestable person)*

Needs word relationships in a context

- *The chair emphasised the need for adult education*

Very common in day to day communications

Satellite Channel Ad: *Watch what you want, when you want* (two senses of watch)

e.g., Ground breaking ceremony/research

A word typically follows -> one sense per discourse

It says that the word when appears in a whole document->it appears in only one sense.

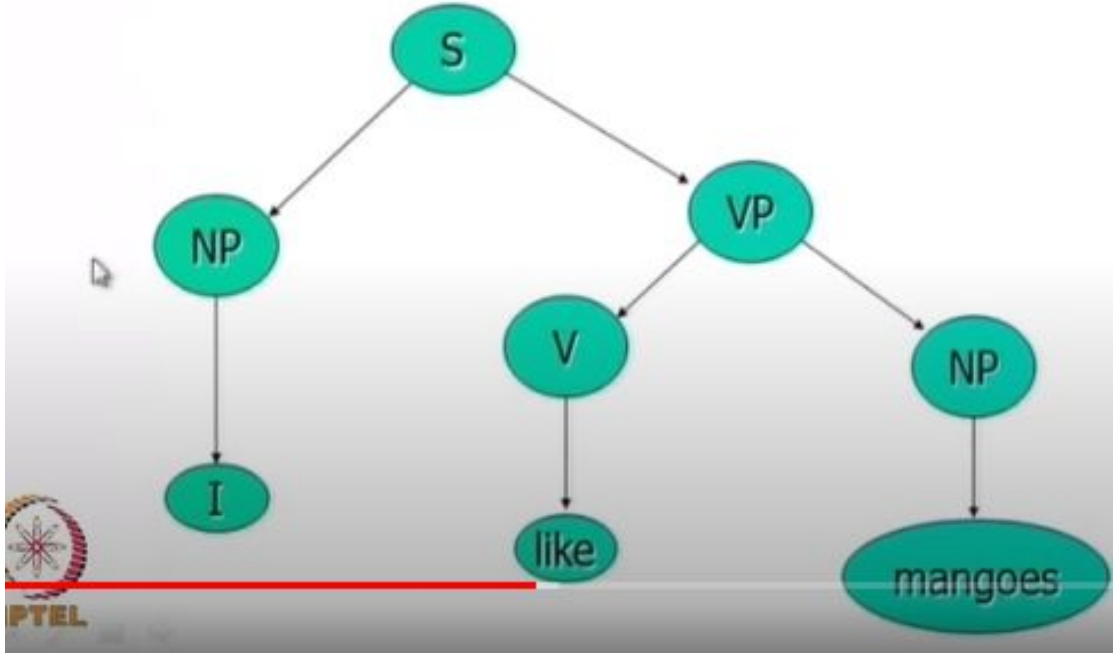
->Predominant sense of dog ->animal

-> In the sentence of chair-> the word emphasised is the cue for disambiguation.

->Computer algorithm needs to be generated on how to obtain that clue from the sentence.

Syntax Processing Stage

Structure Detection



Challenges in Syntactic Processing: Structural Ambiguity

- Scope
  1. *The old men and women were taken to safe locations*
  (old men and women) vs. *((old men) and women)*
  2. *No smoking areas will allow Hookas inside*
- Preposition Phrase Attachment
  - *I saw the boy with a telescope*
    (who has the *telescope?*)
  - I saw the mountain with a telescope
    (world knowledge: *mountain* cannot be an *instrument of seeing*)
  - I saw the boy with the pony-tail
    (world knowledge: *pony-tail* cannot be an *instrument of seeing*)

Challenges in syntatic processing : structural ambiguity

SCOPE

1. The old men and women were taken to safe locations

(old men and women) vs (old men) and women)

2. No smoking areas will allow Hookas inside

Preposition phrase attachment
(possibility of multiple phrase attachment)

e.g. I saw the boy with a telescope

'with a telescope' is attached to what?

e.g. I saw the mountain with a telescope

e.g I saw the boy with pony-tail
(no ambiguity!)

Structural Ambiguity

A camera man shot the man with the gun when he was near Tendulkar.

Ambiguity of a sentence

- Multiple meaning of word
- Multiple attachment point of preposition phrase
- Clause attachment point

Higher level knowledge needed for disambiguation

- Semantics
  - *I saw the boy with a pony tail* (*pony tail* cannot be an instrument of seeing)
- Pragmatics
  - *((old men) and women)* as opposed to *(old men and women)* in "*Old men and women were taken to safe location*", since women- both and young and old- were very likely taken to safe locations
- Discourse:
  - *No smoking areas allow hookas inside, except the one in Hotel Grand.*
  - *No smoking areas allow hookas inside, but not cigars.*

Semantic Analysis

- Representation in terms of
  - Predicate calculus/Semantic Nets/Frames/Conceptual Dependencies and Scripts
- *John gave a book to Mary*
  - Give action: Agent: John, Object: Book, Recipient: Mary
- Challenge: ambiguity in semantic role labeling
  - *(Eng) Visiting aunts can be a nuisance*

Agent ?

Object of visiting ?

Visitor?

Pragmatics

- Very hard problem
- Model user intention
  - *Tourist (in a hurry, checking out of the hotel, motioning to the service boy): Boy, go upstairs and see if my sandals are under the divan. Do not be late. I just have 15 minutes to catch the train.*
  - *Boy (running upstairs and coming back panting): yes sir, they are there.*
- World knowledge
  - *WHY INDIA NEEDS A SECOND OCTOBER (*ToI, 2/10/07)

How a sentence is processed by user?

Modelling the user intention

Discourse

Processing of *sequence* of sentences

Mother to *John*:

John go to school. It is open today. Should you bunk? Father will be very angry.

Ambiguity of *open*

bunk what?

Why will the father be angry?

Complex chain of reasoning and application of world knowledge

Ambiguity of *father*

father as *parent*

or

father as *headmaster*

Complexity of connected text.

How does John know what is the object of bunking ?

World Knowledge - > Disciple important, going to school is important, etc.

John was returning from school dejected. Today was the math test. He couldn’t control the class. Teacher shouldn’t have made him responsible. After all he is just a janitor.