Stages of NLP

Lecture 2

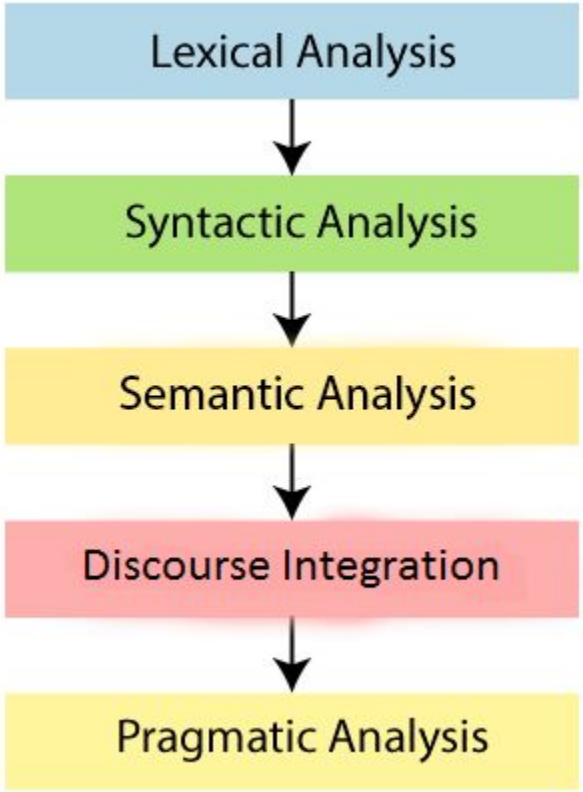
So far

- NLP- Course introduction
- Intuitive introduction
- Formal Definition
- What we do in NLP
- Why NLP is hard
    - Ambiguity

Today: Stages of NLP

Stages of NLP

- Morphology → deals with smallest parts of words that carry meaning
- Lexical → deals with formation of words
- Syntactic → grammar and structure of sentences
- Semantic → meaning of sentence
- Discourse → deals with the structure of different kinds of text.
- Pragmatic → deals with the knowledge that comes from the outside world

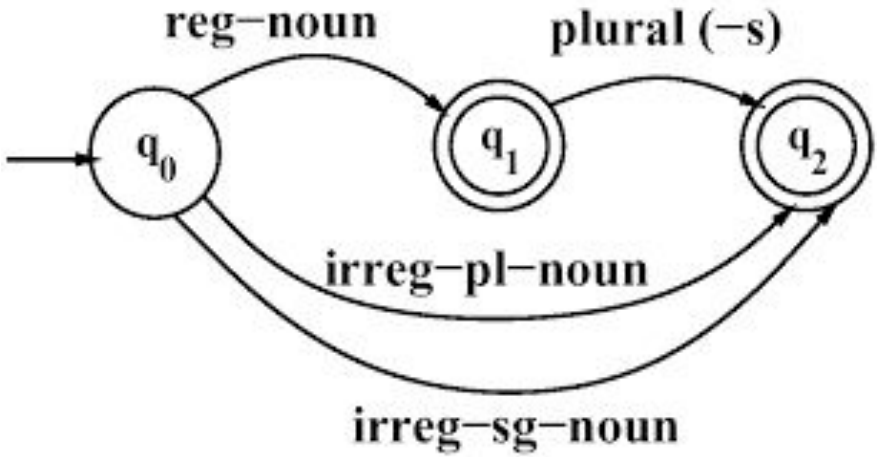


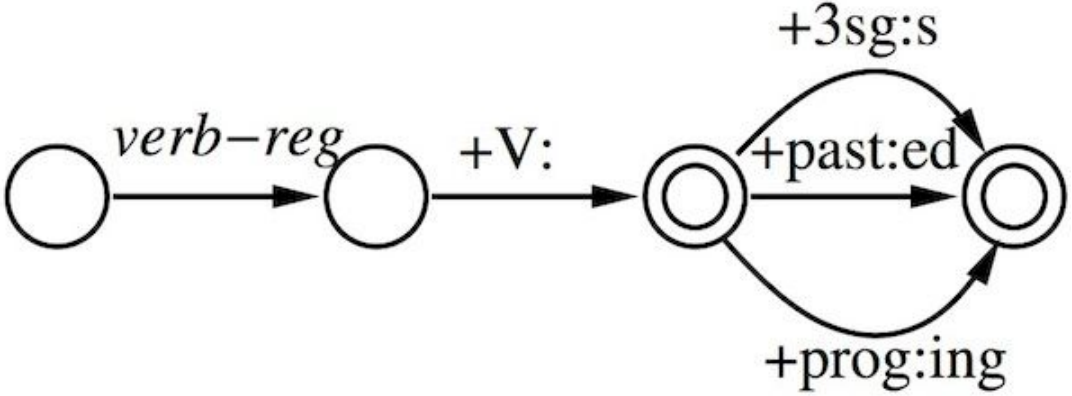
Morphology

- Word formation rules from root
  - Noun→ Plural (boy: boys)
  - Verb → past/past participle/present participle (play:played, playing)
  - Modifiers (un-happy) (suffixes/prefixes)
- First crucial step in NLP
- Dravidian Languages have rich morphology
- Computational Problems:
  - **How to represent word formation rules?**
  - **How to effectively find morphemes from the word?**
- **Finite state machines**

Morphology

Example





Lexical Analysis

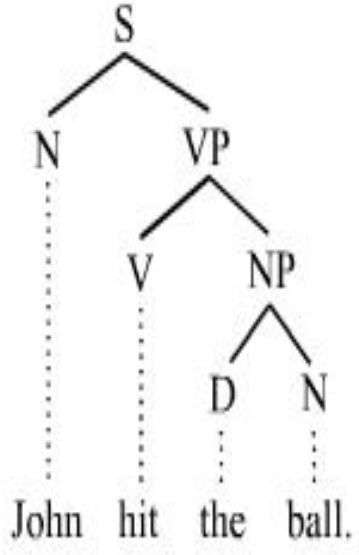
- Identifying the word and its properties
- Eg: Dog (Noun, take ‘s’ in plural form, animate, 4 leg, carnivore)
- Many words take multiple meaning. → Ambiguity
- Disambiguating the right sense is a challenge (WSD)
- How to deal with WSD?

Lexical Disambiguation

- Part of speech disambiguation
  - Dog (as noun)
  - Dog (as verb) (dogging/dogged)
- Sense Disambiguation
  - Dog (as animal)
  - Dog (as very detestable person)
- Needs word relationship in a context
  - Eg: The *Chair* emphasised the need for adult education
  - Ground breaking ceremony (think!!)
- Note: Each word a have predominant meaning, how to model it?
- Also new words and their usages are coming.. (googling, xeroxed, digital trace, texto, speako (like typo)…)

Syntax Analysis

- Give a structure to the sentence
- Structure → Parse Tree
- Grammars plays an important role
- Challenge: Syntactic Ambiguity



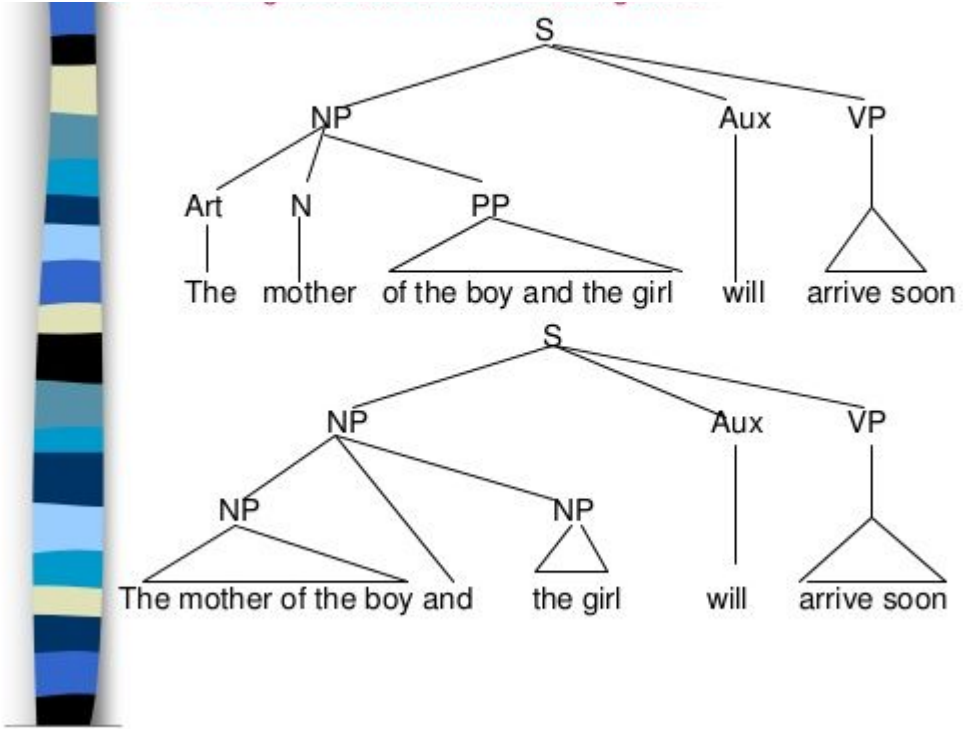
Constituency-based parse tree

Syntax Level Ambiguity

I ate dosa with chatni

- I ate dosa wither broth

Solution Probabilistic grammar



Some more challenges in Syntax level

Scope (of adjectives/clause→ attachments):

- Old men and women were taken to safe location
  - (old men and women) or (old men) and women
- No smoking areas will allow hookah inside
  - No→ not any(quantifier, like all, any, some) No(smoking areas……. inside)
  - (No smoking areas) will allow hookah inside

Preposition attachment

- I saw the boy with a telescope
- I saw the mountain with a telescope
- This kind of ambiguity is common in Newspaper headlines
  - Teacher Strikes Idle Kids
  - Squad Helps Dog Bite Victim (more examples :https://www.departments.bucknell.edu/linguistics/synhead.html)

Ambiguity arises

- Word with multiple meaning
- Multiple preposition attachments
- Clause attachment

Higher level knowledge required for disambiguation

Semantics

- I saw a boy with a ponytail (from word knowledge *ponytail cannot be an instrument of seeing*)
    - I saw a boy with a ponytail)

Pragmatism

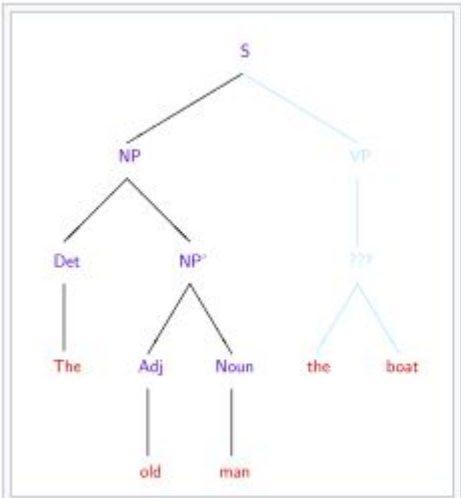
- Old men and women were taken to safe location
    - (Old men) and women…. Will be more practical, since both old and young women were very likely taken to safe location

Discourse

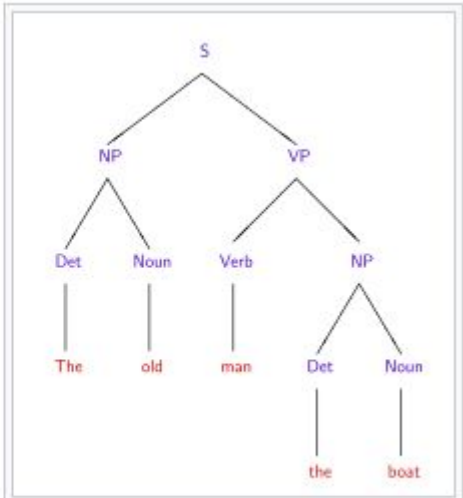
- No smoking area allow hookah inside, ***except the one in Hotel Grand***
- No smoking area allow hookah inside, *but not cigars*

Challenges in Parsing: Gardenpath

- A garden-path sentence is a grammatically correct sentence that starts in such a way that a reader's most likely interpretation will be incorrect;
- *“ The old man the boat “*
- “*The complex houses married and single soldiers and their families.”*



Suggesting uncompletable parse



Less obvious complete parse

Semantic Analysis

- Represent the meaning in logical level (use predicate calculus)
  - Semantic Nets (graphs), Frames (slots and fillers), Conceptual Dependendencies, Scripts (templates/stereotypes)
- Ex: John gave a book to Mary
  - Action: gave
  - Agent: John
  - Object: book
  - Recipient: Mary
  - Gave(john, book, mary)
- Semantic Role Labeling (SRL)
- Karaka in Sanskrit (have a look)

Ambiguity in SRL

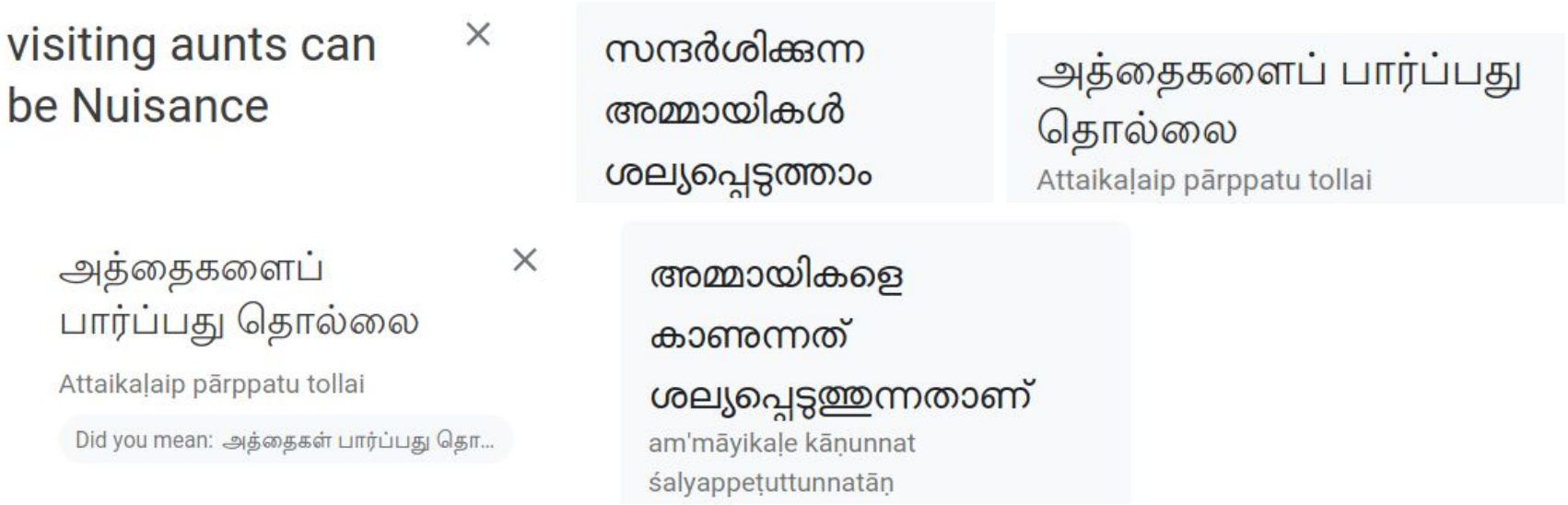
Visiting aunts can be a nuisance

```
>>> text14 = nltk.word_tokenize("Visiting aunts can be a nuisance")
>>> nltk.pos_tag(text14)
[('Visiting', 'VBG'), ('aunts', 'NNS'), ('can', 'MD'), ('be', 'VB'), ('a',
'DT'), ('nuisance', 'NN')]
```

- **Another alternative POS tagging**

```
[('Visiting', 'JJ'), ('aunts', 'NNS'), ('can', 'MD'), ('be', 'VB'), ('a',
'DT'), ('nuisance', 'NN')]
```

Ambiguity in SRL



Pragmatics

- 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.*

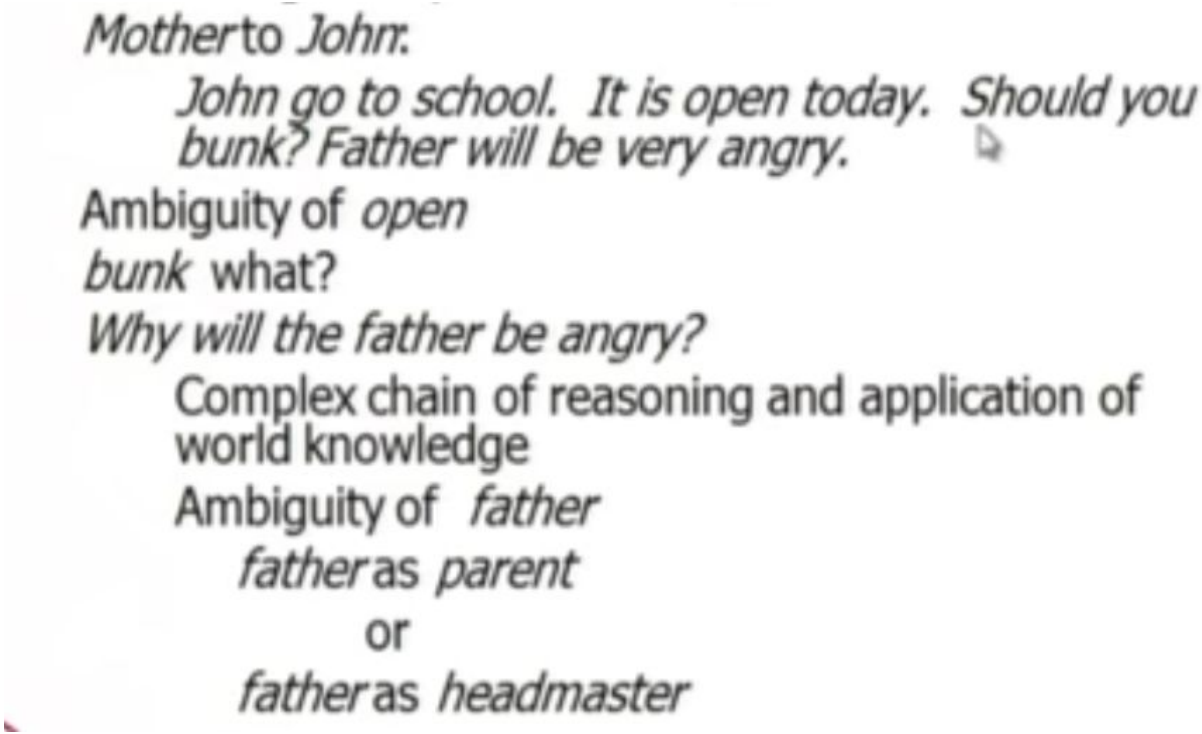
Pragmatics

- World Knowledge is required
- Eg: “WHY INDIA NEED A SECOND OCTOBER” (TOI)

Should know:

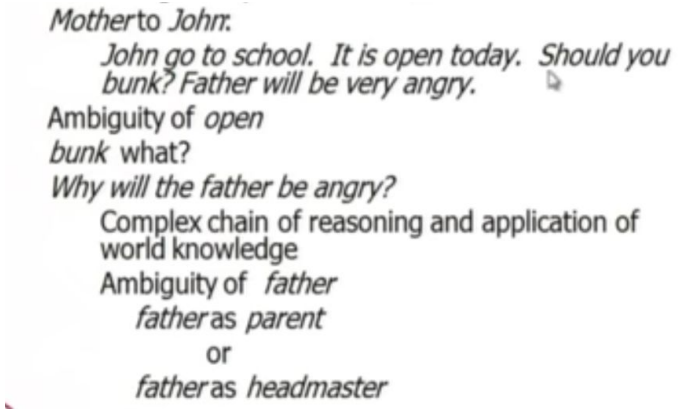
- People interest
- World knowledge
- Context knowledge
- Common sense
- Model conversations

Discourse

- Processing of sequence of sentences
- 

Discourse

- Sequence of sentences
- Should memorize meaning of previous sentence
- Should predict meaning of future sentence



Two Approaches

- Classical NLP (traditional/ rule based)
- Statistical /Machine Learning based