**Stages of Processing**

* Phonetics and phonology – Processing of sounds
* Morphology – Processing of word forms
* Lexical Analysis – Storage of words in a dictionary
* Syntactic Analysis – Understanding the syntactic structure of the sentence
* Semantic Analysis – understanding the meaning of the sentence
* Pragmatics – Intention of person in saying the sentence
* Discourse – Relation between multiple sentences

**Phonetics:**

* **Processing of speech**
* Challenges of phonetics

-Homophones: bank(finance) vs bank(river)

-Near Homophones: maatraa vs. maatra (hindi word)

-Word Boundry:

I got [u/a]plate

-Disfluency: ah, um, hmm, ahem etc.

**Morphology:**

* **Word formations 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.

**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)

Challenges: Lexical or word sense disambiguation

First step: part of speech disambiguation

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

Second step: Sense Disambiguation

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

**Syntax Processing Stage**

* Structure Detection:
  + I like mangoes

Noun Phrase<- Sentence -> Verb Phrase

Noun Phrase -> I Verb <-Verb Phrase-> NounPhrase

V-> like NP->mangoes

**Syntax Processing Stage contd..**

* **Scope**

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

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

1. No smoking areas will allow Hookahs inside.

* Preposition Phrase Attachment
* I saw the boy with a telescope.

(who was the telescope?)

* I saw the boy with the pony-tail

(world knowledge: pony-tail cannot be an instrument of seeing)

**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
* Challenges: ambiguity in semantic role labelling
* (English) Visiting aunts can be a nuisance

**Pragmatics**

* Very hard problem – checking intentions of sentences used.
* Model user intention
* Tourist (in a hurry, checking out of the hotel, monitoring to service boy): Boy, go upstairs and see if my shoes are under the bed. 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.
* Machines do not have very good understanding of sentences.

**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.

1. Ambiguity of open. What is open? Windows or door? Should it not be school is working today.
2. Bunk what? – Subject is not defined.
3. Why will the father be angry?

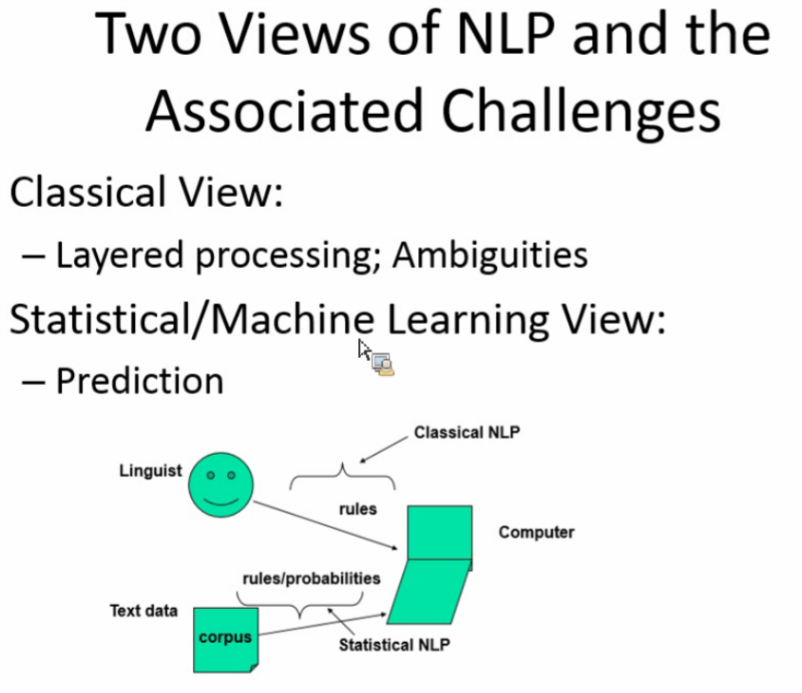
-Complex chain of reasoning and application of world knowledge.

-Ambiguity of father

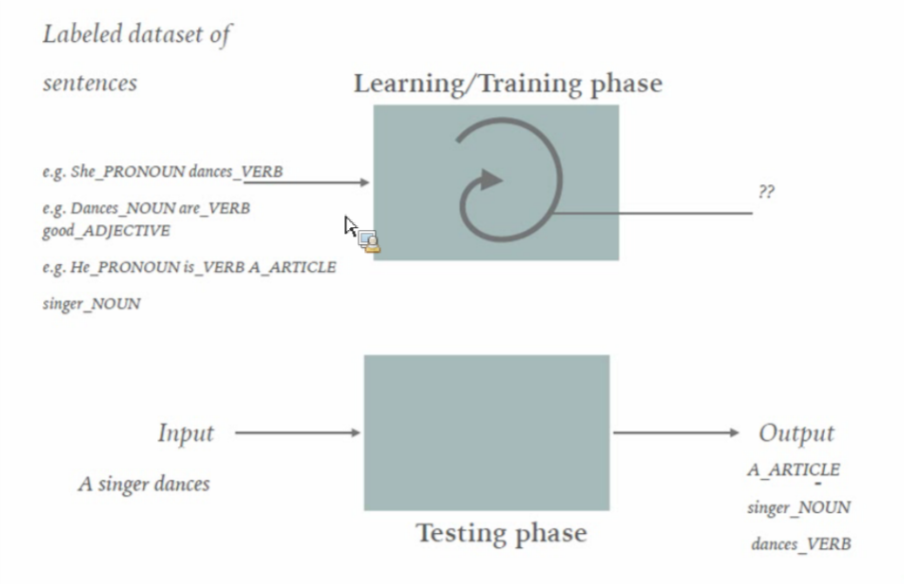
Father as parent

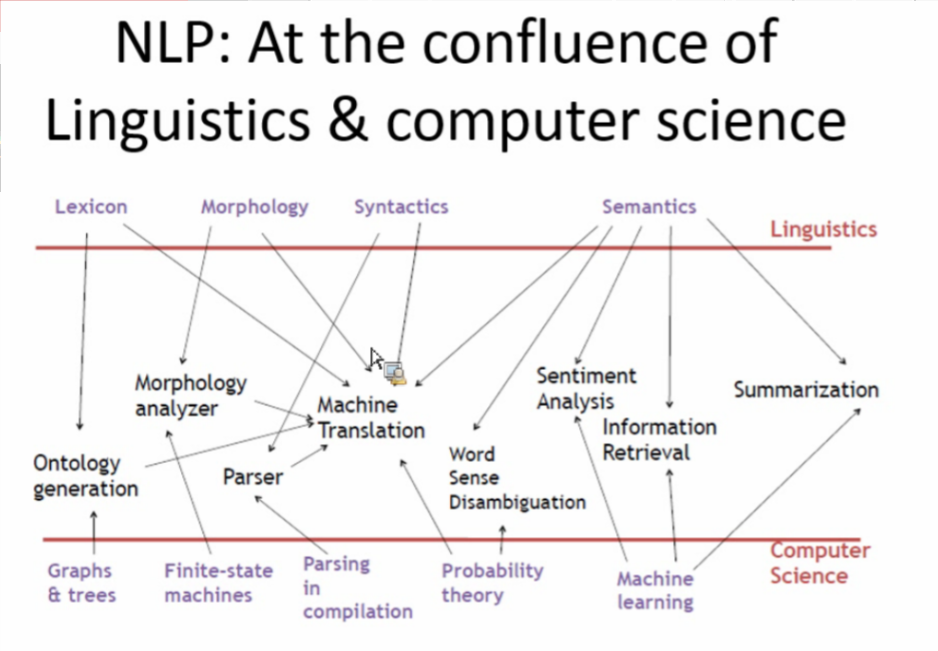
Or

Father as headmaster of school



**Structure of Statistical/ Machine Learning View:**

****

****

**Lexical resources:**

**Corpus**

* A collection of text, it can be a plain text document containing data or a structured XML file containing data along with its descriptors and headers.
* Corpora – Plural form of corpus
* Examples:
* Monolingual Corpa
* Comparable
* Parallel
* Annotated corpora: pos-taggend, sense- tagged.
* Monolingual corpus:

e.g. A boy is sitting in the kitchen .

A boy is playing tennis.

* Comparable Corpus

-collection of “similar” texts in different languages or in different varieties of a languages.

- A pair of corpora in two different languages, which come from the same domain.

- An example would be a corpus of articles about football from English and Danish newspaper; or legal contracts in Spanish and Greek.

* Parallel Corpus:

-The text in one language are translations of texts in the other languages.

* Annotated Copus:

-Corpus annotated with different label to each unit in the corpus.

* Pos-tagged:

-She\_PRON eats\_VERB rice\_NOUN

* Sense-tagged corpus

-The city\_18406385 is\_22579744 famous\_539872

**What is wordnet?**

Dictionary = Words + meanings

Wordnet = word + meaning + semantics and lexical relations

* A lexical knowledge database for a language
* Consists of synsets and lexico-semantic relations
* Categorizes synsets into four main parts of speech categories:
  + Noun, adjectives, adverbs and verbs
* Monolingual Wordnet
  + - English, Hindi, Chinese
  + Multilingual WordNet
    - IndoWordNet, EuroWordNet, BabelNet

Link: <http://wordnetweb.princeton.edu/perl/webwn>

WordNet Lexico-Semantic Relations

* Synonymy: relation between word having synset.

{plant, flora} ‘plant’ and ‘flora’ are related through synonymy relation.

* Antonym: Relation between words having opposite meaning.

‘day’ and ‘night’

* Gradation:

‘morning’, ’afternoon’, ’evening’ are related through gradation relation

* Hypernym/Hyponym: is a kind of relation where

‘fruit’ is a hypernym of ‘mango’ and ‘mango’ is a hyponym of ‘fruit’.

* Meronymy/ Holonymy: part- whole relation

‘hand’ is a meronym of ‘body’ and ‘body is a holonym’ of ‘hand’

* Entailment:

‘snore’ entails ‘sleep’

* Attribute: relationship between noun and adjective synsets.

‘hot’ is a value of or attribute of ‘temperature’

* Nominalization: Relation between noun and verb synsets.

‘service’ nominalizes the verb ‘serve’

* Ability Link: specifies the inherited features of a nominal concepts.

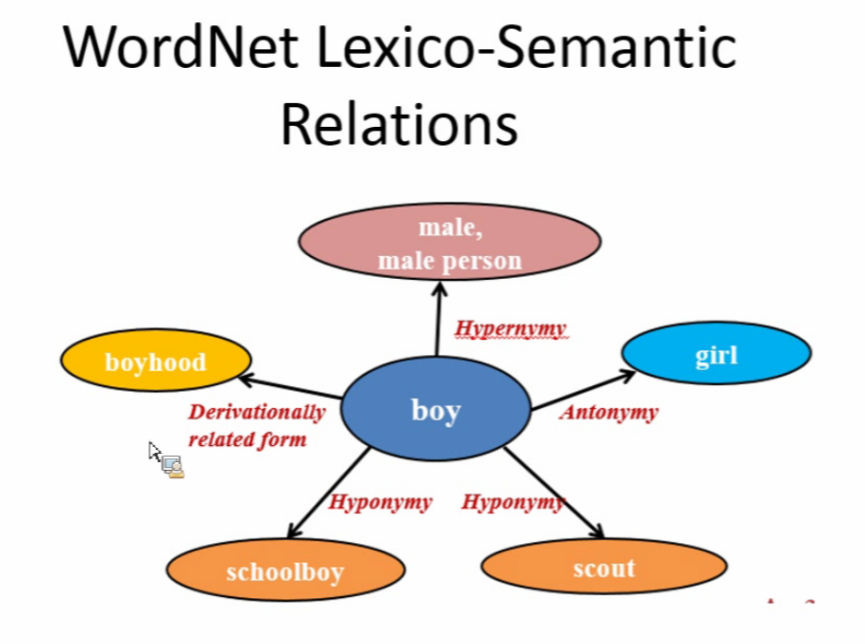
‘animal’ and ‘walk’, ‘fish’, ‘swim’

* Capability Link: relationship specifies the acquired features of a nominal concept.

‘person’ and ‘swim’

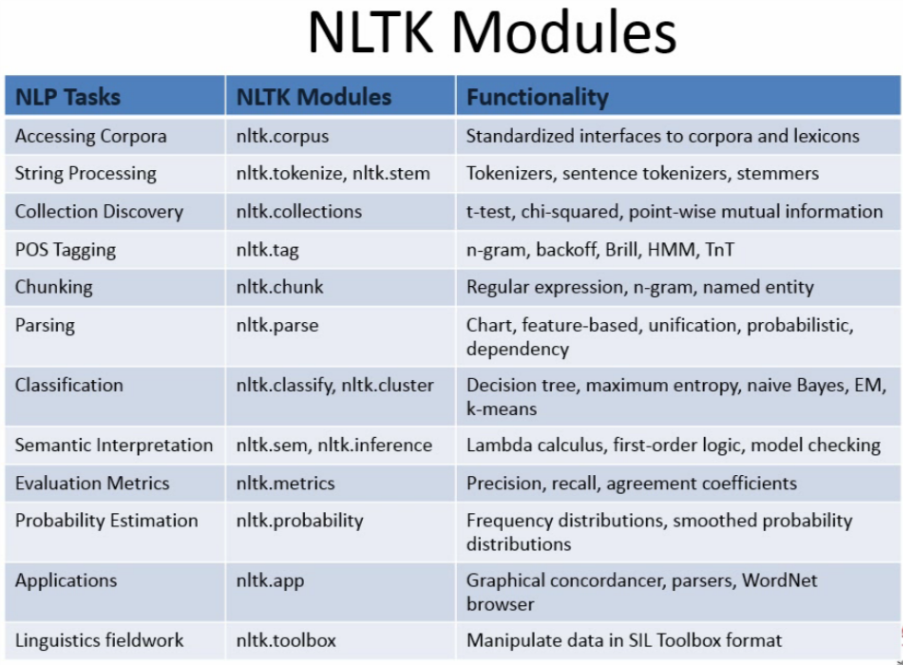
* Functional Link: relationship specifies the function of a nominal concept

‘vehicle’ and ‘move’ and ‘teacher’ and ‘teach’



**Modules Used in NLP**

**NLTK:**

****