→ In HMM, the observation - vectors of spectral features whereas in markov model, it is simply the phones

ASSIGNMENT - I

22/09/22.

Q. Write down the POS tags for the following sentences

1. I prefer a morning flight.
   PP VB DT NN NN

2. Show me the cheapest flight fare.
   VB PP DT RJJS NN NN

3. Does this flight serve dinner?
   VBZ DT NN VB NN
   ?

4. Which flight serves dinner?
   WDT NN VBZ NN

5. Which book do you have in your classroom?
   WDT NN VB PP VBP IN PPZ NN

6. Which book do you borrow from the library?
   WDT NN VB PP VB IN DT NN

7. Can you give some information about Paris?
   MD PP VB (me) PP DT NN IN NNP

8. How can I go from Chennai to Trichy?
   ? MD PP VB IN NNP IN NNP

9. I told Harry to go to London.
   PP VBD NNP IN VB IN NNP

10. How does the flight-543 arrive in Atlanta?
    ? VBZ DT NNP VB IN NNP
    ?

11. Flight-208 will arrive at 3'o clock.

NNP VB IN

12. A bank can hold the investments in a custodical account in the client's name.

DT NN MD VB DT NNS IN DT JJ? NN IN

DT ? NN

Q. Draw the pronunciation networks for the following words using simplified markov model and HMM.

1. ate
2. good
3. return
4. believe
5. flight.

Syntax analysis

25/09/2022

→ Relationship between words in the sentence

→ Pos relation between various classes of words in Grammar

e.g.:-
- Nouns — name
- Verbs — eat, like, go
- Auxillaries — Auxillary verbs represent tense of actions
  - Past, present, future
  - will, would
- Adjectives
- Prepositions
- Determiners
- Pronouns
  - Personal — he, she, it, they
  - Possessive — him, his, her, their, my

e.g.:- london — Proper nouns

e.g.:- Garden City — Common nouns

Adverbs

locative	temporal	degree	manner
down the lane, here, there	to day, yesterday evening, 11 o clock	-degree of action very	slowly,

Adjectives

age – young / old
speed - slow / fast (Ram walks fastly)

Prepositions

on, at, under
The book is on the table.

Determiners

- ~~succeeded~~ precede noun
  e.g. – a, the, an, this, that
  - A book, The book, That thing

Nouns (also splitted as)

count nouns	mass nouns
- has plural, can be counted goat - goats boat - boats	- cannot be counted but represents homogeneous group snow, salt, stars in the sky

Ambiguity in different phases of NLP

1. **Morphology**
   - Word with same PoS, different meaning

e.g.:- Ram deposited his money in Central Bank → financial bank
(N)

There is a river bank in Paris → river bank
(N)

- same word, different PoS

e.g.:1) Ram created issues.
(N)

He issues certificates
(V)

2) Book that flight
(V)

The book is on the table.
(N)

2. Syntax analysis

- more than one parse tree for the same sentence.

A = A * B + C

A = A * B + C | A = A * B + C

e.g:- Teacher strikes idle kids — ambiguous
N V Adj N

3. Attachment ambiguity

- If a constituent fits more than one place at the sentence

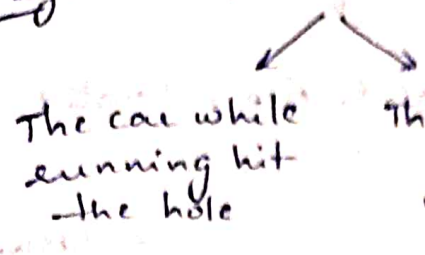
John saw Mary with a telescope.

John saw Mary with a telescope.

4. Semantic ambiguity

- A sentence can be interpreted in more than one form

eg:- The car hits the hole while it was moving



5. Pragmatics

You are late

↓

What about the intention of the sentence? — informing / ~~criticizing~~

6. Discourse

We gave the monkeys the bananas because they were hungry.

We gave the monkeys the bananas because they were overripe.

monkeys

bananas.

GRAMMAR

Grammar is a four tuple

1) a start symbol S
2) a set of non-terminals — generalization of terminals
3) a set of terminals — words
4) a set of production rules in the form

$\alpha \rightarrow \beta$ α: NT

β: Terminal / NT

Grammatical sentences:-

- sentences derived by a CFG for a language defined by that grammar.

Non-grammatical sentences:-

- Not derived by a CFG for L — not defined by a formal grammar.

A language can be recognized by

- grammar
- FSA/Transducer } morphological phase.
- RE

If a sentence $\in L(G)$ then sentence is valid, else not valid.

27/09/22

Types of sentences (that support syntax analysis)

1. **Declarative**

$S \rightarrow NP\ VP$

e.g.:- I (NP) prefer a morning flight (VP).

2. **Imperative**

- actions / commands / suggestions
- begins with a verb

$S \rightarrow VP$

e.g.:- Show me the cheapest flight fare. (VP)

3. **Yes/No**

- begins with an auxillary verb followed by NP, VP

$S \rightarrow Aux\ NP\ VP$

e.g.:- Does (Aux) this flight (NP) serve dinner (VP)? → do / does / can / would / will / could, etc.

4. **Wh-questions**

- begins with who, why, where, which, how?

1. **wh-subject questions**

- Similar to declarative

$S \rightarrow wh\text{-}NP\ VP$

e.g.:- Which (wh) flight (NP) serves dinner (VP)?

2. Wh-non subject

- similar to yes/no
- preceded by wh-NP

$S \rightarrow$ Wh-NP Aux NP VP

e.g.:- Which book do you borrow from library?
Wh NP Aux NP VP

Q. Consider the grammar G defined ~~the~~ for language L.

$S \rightarrow NP\ VP$

$NP \rightarrow Det\ Nominal \mid Noun$

$Nominal \rightarrow Adv\ Nominal \mid Noun$

$VP \rightarrow V\ NP$

$PP \rightarrow P\ NP$
(Prepositional phrase)

e.g.:- I prefer a morning flight. (find if $S \in L(G)$ or not).

$S \rightarrow NP\ VP$

$N\ VP$

$N\ V\ NP$

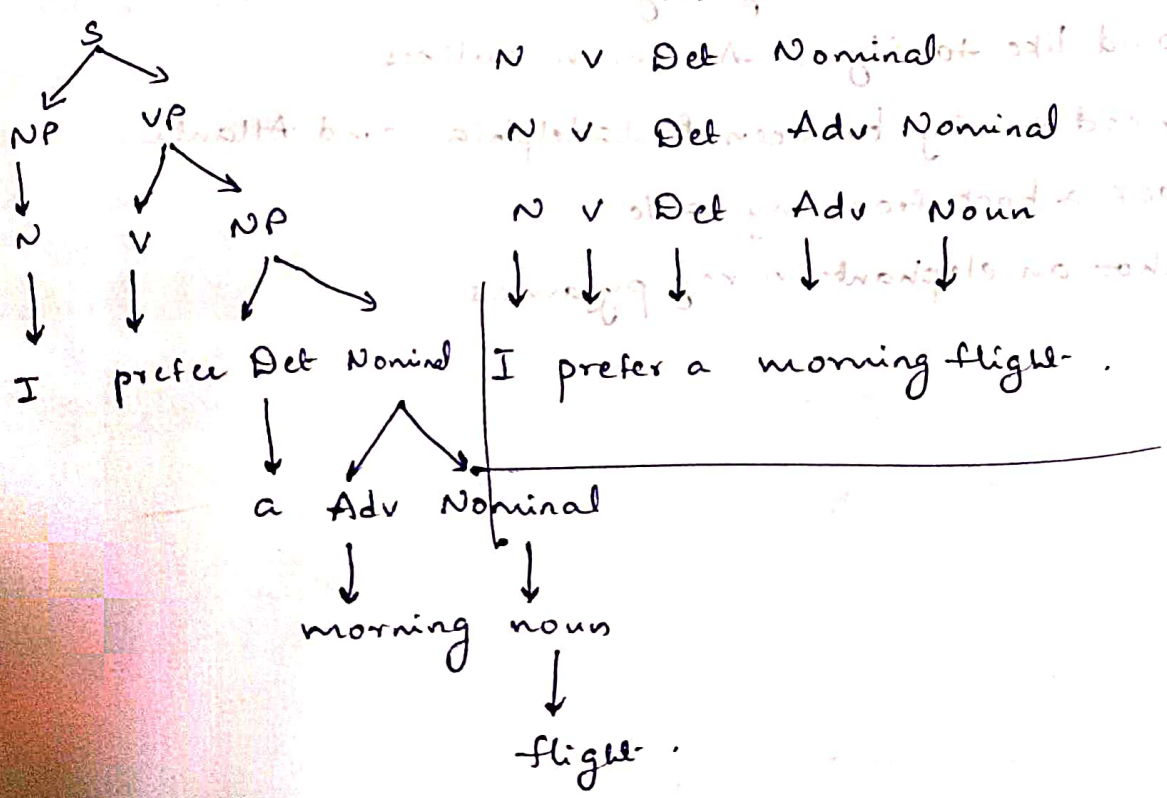
$N\ V\ Det\ Nominal$

$N\ V\ Det\ Adv\ Nominal$

$N\ V\ Det\ Adv\ Noun$

I prefer a morning flight.

Parse tree



Q. Consider the grammar G defined for a language L

$S \rightarrow NP\ VP$

$S \rightarrow Aux\ NP\ VP$

$S \rightarrow VP$

$NP \rightarrow Det\ Nominal$

$Nominal \rightarrow Noun$

$Nominal \rightarrow Noun\ Nominal$

$NP \rightarrow Noun$

$VP \rightarrow Verb$

$VP \rightarrow Verb\ NP$

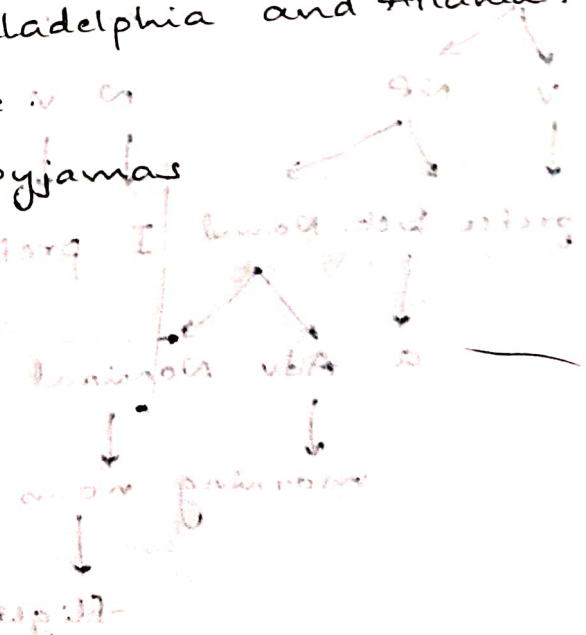
$NP \rightarrow NP\ PP$

$PP \rightarrow P\ NP$

$NP \rightarrow Possessive\ Noun\ Noun$.

Check whether the following sentences $\in L(G)$ or not.

1. Does this flight include a meal?
2. John prefers a morning flight.
3. I would like to fly on American airlines.
4. Ib need to fly between Philadelphia and Atlanta.
5. I took a book from my table.
6. I shot an elephant in my pyjamas



29/09/22

Q. Construct the parse trees for the following sentences

1. I prefer a morning flight
2. show me the cheapest flight fare
3. Does this flight serve dinner?
4. Which flight serves dinner?
5. which book do you have in your classroom?
6. which book do you borrow from the library?

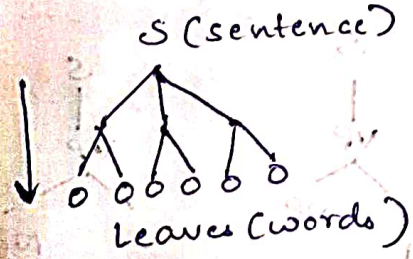
Q. Derive the grammar and construct the parse tree from the following sentences.

1. Can you give me information about Paris?
2. How can I go from Chennai to Trichy?
3. I told Harry to go London.
4. I want to go to eat some Italian food today.
5. ~~Why~~ When does flight 573 arrive in Atlanta?
6. flight 208 will arrive at 3'0 clock.
7. A bank can hold the investments in a custodial account in the client's name.

10/10/22

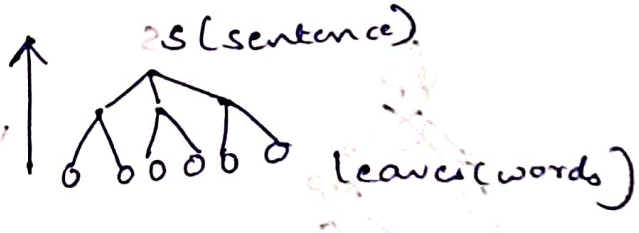
Types of parsing

1. Top-Down parsing



- Starting from root (S - sentence) down towards the leaves (words)

2) Bottom-up parsing



- from leaves (words) up towards the root - S (sentence)

- never wastes time in generating the sub-trees (that are not part of S-rooted tree)	- subtrees that have no hope of leading to S are generated in wild abandon

Q. Consider the grammar G defined for a language L.

S → NP VP

S → Aux NP VP

S → VP

NP → Det Nominal

Nominal → Noun

Nominal → Noun Nominal

NP → Proper-noun

VP → verb

VP → verb NP

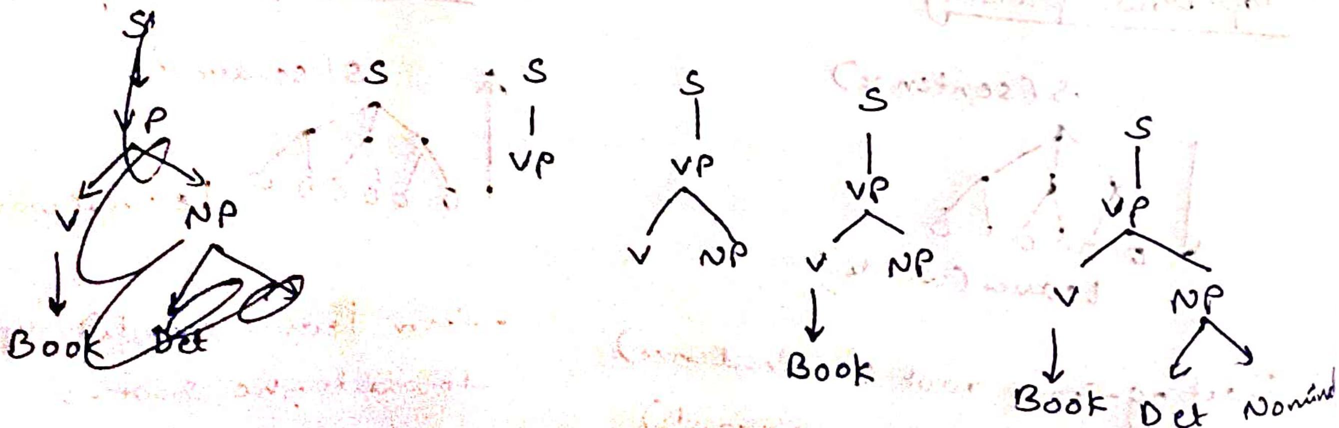
NP → NP PP

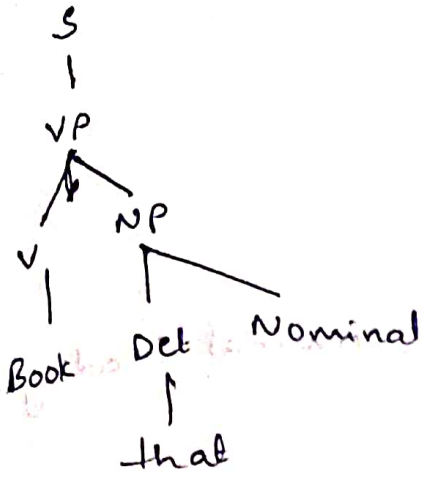
PP → P NP

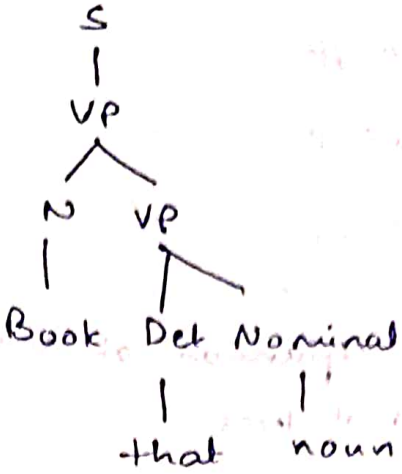
NP → Poss Noun Noun

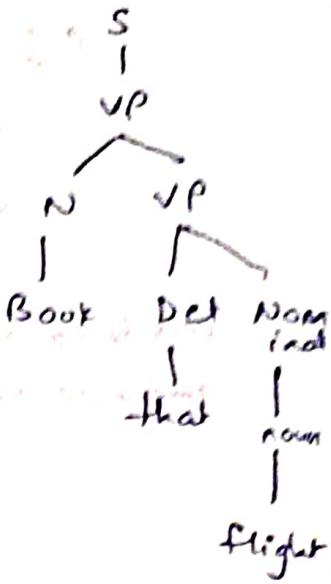
S: Book that flight

<u>Top-Down parsing</u>

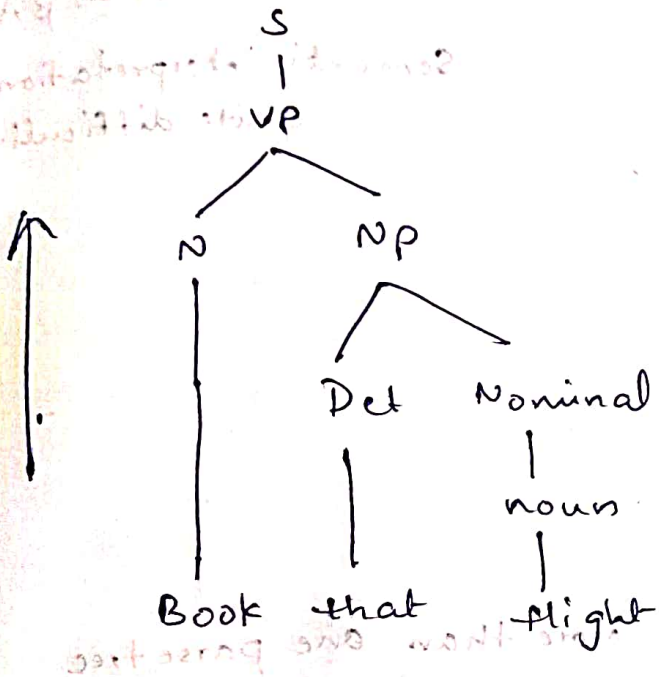








Bottom-up parsing



Problems in top-down parsing

1) Left recursion
2) Structural Ambiguity
3) Repeated parsing of subtrees.

1) Left recursion

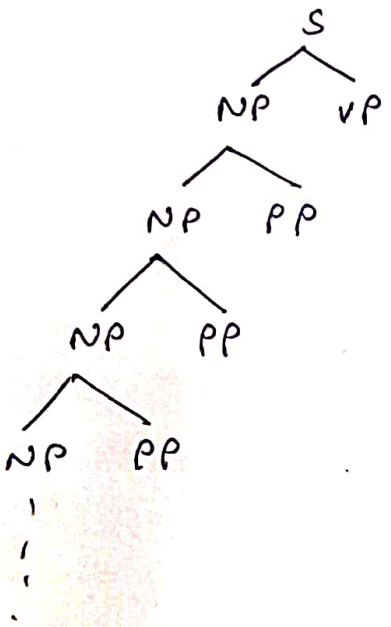
A grammar is left recursive if it contains atleast one non-terminal A such that $A \Rightarrow A\beta | \alpha$

e.g.:- S → NP VP

NP → NP PP

VP → VP PP

- leads to recursive expansion of some non-terminal leading to infinite expansion of trees.

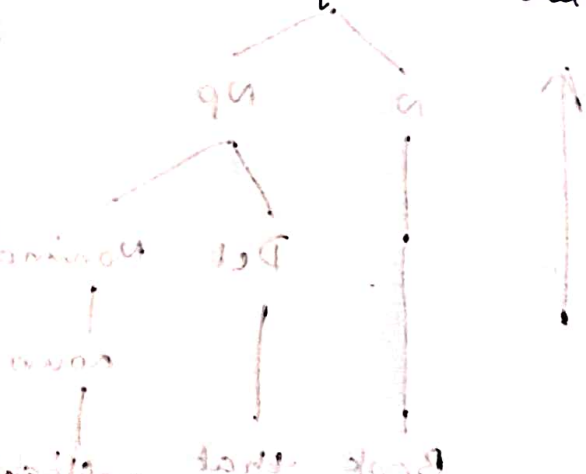


Elimination

$A \Rightarrow A\beta \mid \alpha \Rightarrow A \rightarrow \alpha A'$

$A' \rightarrow \beta A' \mid \varepsilon$

Semantic interpretation - quite difficult



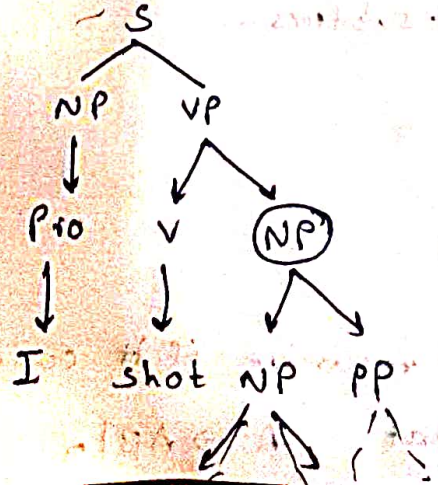
2) Structural ambiguity

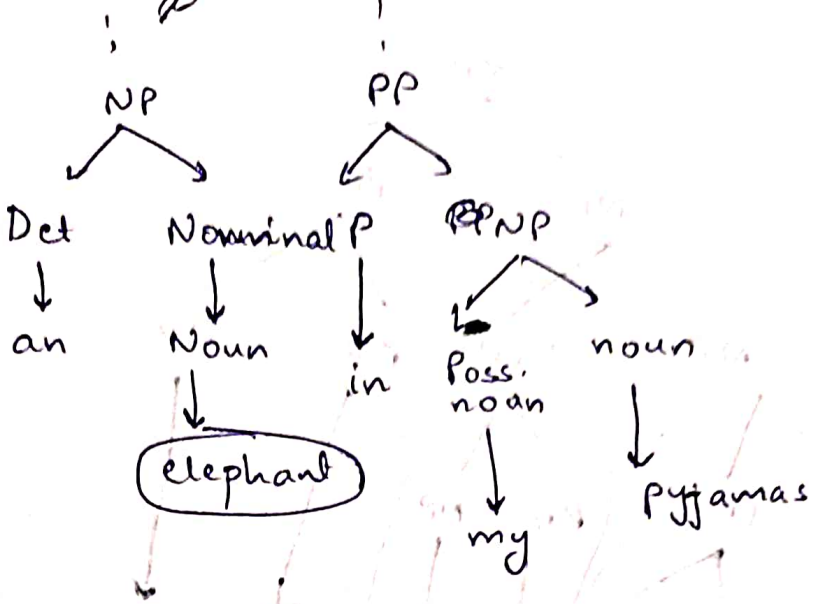
- A grammar that produces more than one parse tree for the same sentence.

e.g.:- I shot an elephant in my pyjamas

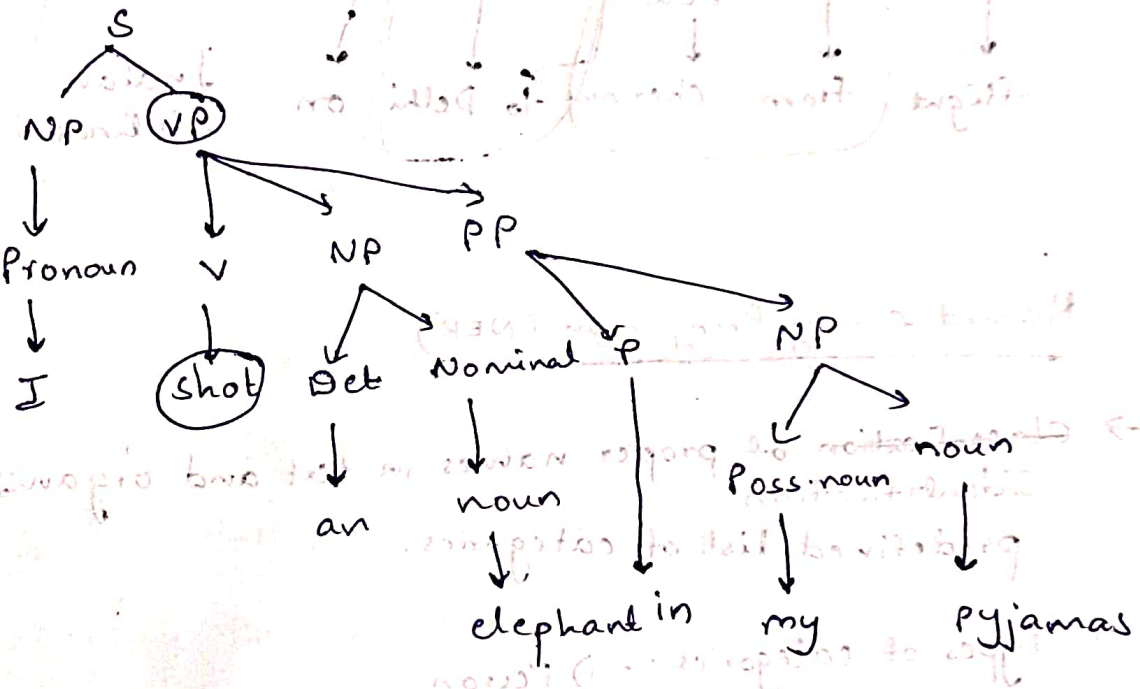
- 2 parse trees - generated for this ambiguous sentence.

i) the phrase in my pyjamas - part of NP headed by elephant





ii) VP- headed by shot



Some other examples for structural ambiguity are:

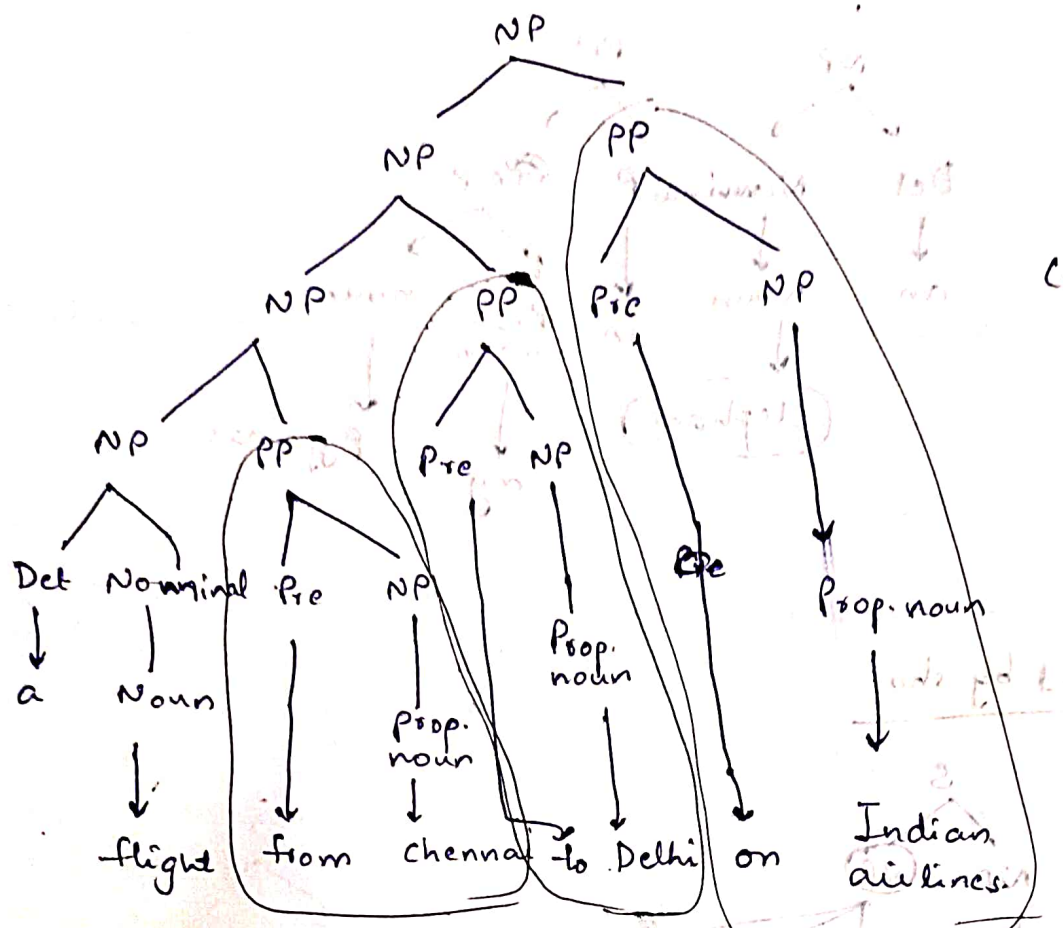
1. Can you book two flights
2. Teacher strikes idle kids.

3) Repeated parsing of sub-trees

Consider the NP →

a flight from Chennai to Delhi on Indian Airlines.

NP PP PP PP



(repeated Parsing of same structure of subtrees)

Named Entity Recognition (NER)

11/10/22

→ ~~Classification~~ Identification of proper names in text and organized into predefined list of categories.

→ Types of categories :- 1) Person
2) Organization
3) Location

→ Key part of IE and IR

Problems

Person vs Artifact

The Ham Sandwich waits for a bill.

Bring me a Ham Sandwich

England won the world cup.

The world cup match took place in England.

Company vs Artifact

Shares in MTV watching MTV

Ambiguity in NE types

JohnSmith → person vs company

Washington → person vs location

May → person vs month

1945 → Date vs Time

Solutions

- task definitions must be clearly specified at the outset.
- simplistic approach for disregarding metonymy.
  e.g.:- England - always identified as a location.
- for each category, there must be a list of examples, counter examples, logic behind the intuition, guidelines.

Approaches

- List look up
- Shallow parsing

1) List look up :- System that recognizes only ~~satisfies~~ entities stored in its lists (Gazettes)

adv - simple, fast, language independent

disadv - maintenance of large lists, can't resolve ambiguity

2) Shallow Parsing

Internal evidence → names often have internal structure

Capword + {city, forest, ...}

e.g.:- mangrove forest
London city

Capword + {street, Road, Avenue}

e.g.:- Gandhi Road
7th Avenue
Market street

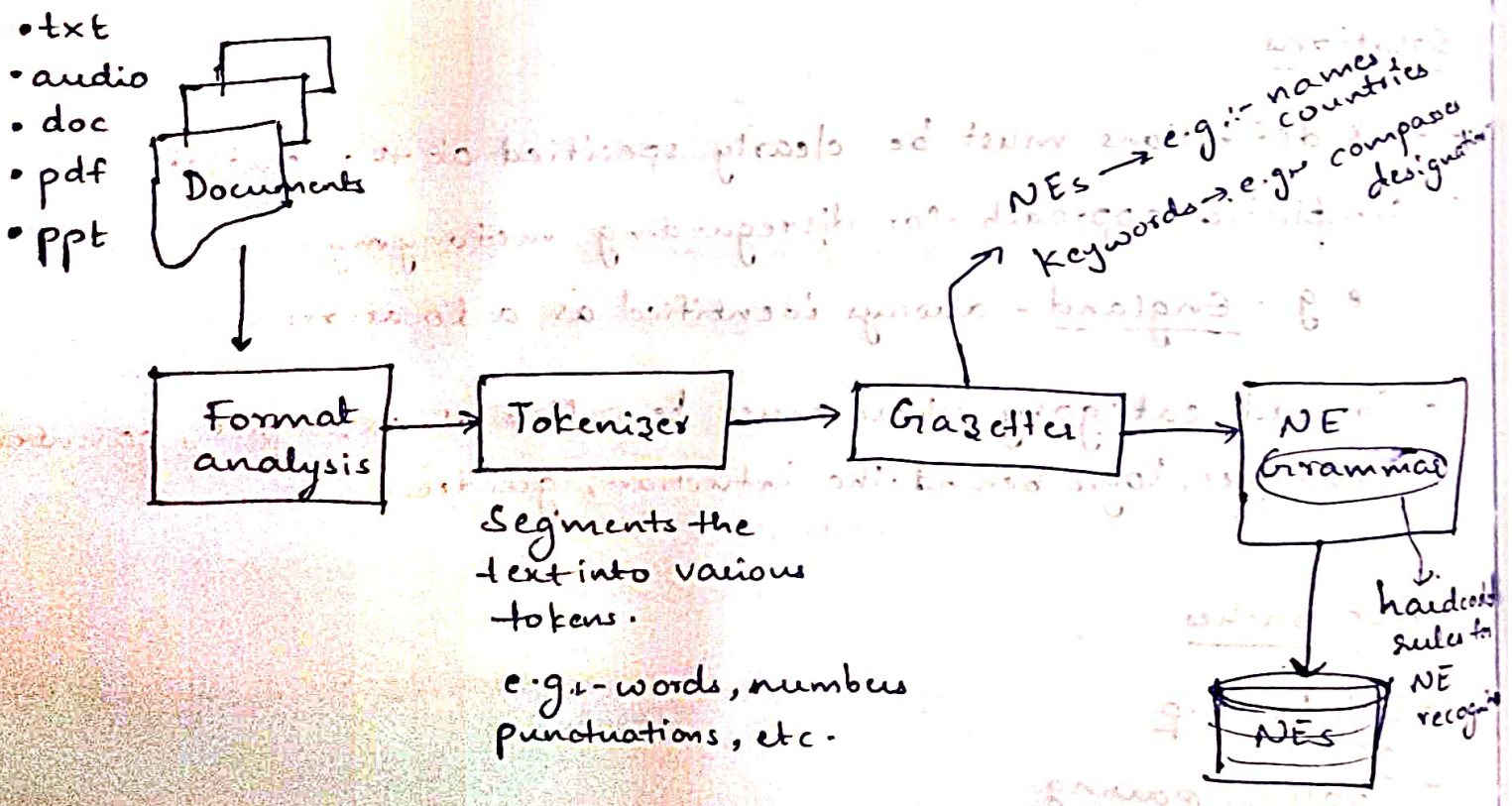
External evidence → into the compass of capword direction

e.g.:- into the south of Africa

→ capword is a Adj Geoword

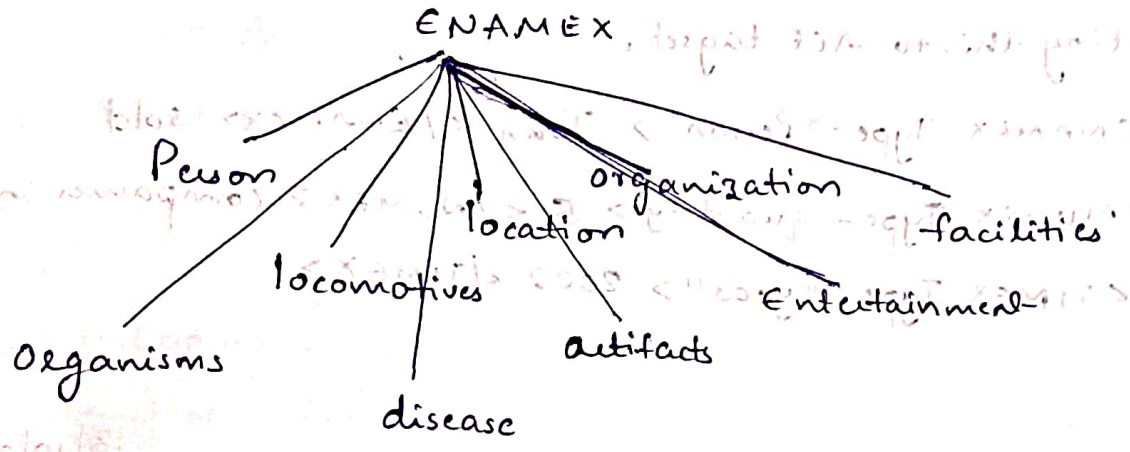
e.g.:- Chennai is a friendly city

NE system architecture



NE hierarchy

1. Entity name (ENAMEX)
2. Numerical expressions (NUMEX)
3. Temporal expressions (TIMEX)



NUMEX

- Distance (km, m, feet, cm, mm)
- Quantity (litres, g, kg, tonnes)
- Money (Rs, $, Euro.)
- Count
  e.g.:- 6 apples, 4 pencils

TIMEX

- Time
  am, pm, morning, evening
- Month (Jan-Dec)
- Date (1st Jan)
- Year (1995, 2002)
- Day (SUN-SAT)
- Special Day (Independence day, Republic day)
- Period (10th century BC, ...)

Representation

- ACE tagset – HTML tags.
  - Automatic content extraction

e.g.:-

John sold 5 companies in 2002.

Converting this to ACE tagset,

<ENAMEX Type="Person"> John </ENAMEX> sold

<NUMEX Type="quantity"> 5 </NUMEX> companies in

<TIMEX Type="year"> 2002 </TIMEX>

13|10|2022

Shallow / Chunk / Partial Parsing

- Divide a sentence into a sequence of chunks.
- Find major but unembedded constituents like NPs, VPs, PPs

Chunks

- non-overlapping regions of a text
- non recursive
  - → a chunk can't contain other chunks
- non-exhaustive
  - → not all words are included in chunks
- doesnot cross the boundaries of major constituents.

Types

i) NP- Noun phrase chunking

[I saw] [a tall man] [in the park]

NP NP NP

2) VP chunking

I saw the man who was older in the garden.
VP VP

3) Question Answering

Which [Spanish explorer] discovered [The Mississippi River]?

4) Prosodic chunking

Study of tunes of rhythm of speech

Applications

- Bootstrapping a more complete parser
- Constructing a tree bank annotated text which other applications can use.
- Extraction of specialized term or multi-words. (medical)
- Information Retrieval

Unchunking

→ Apply Pos tagging to the sentence.

→ Apply the chunk rule
NP or VP or PP

→ Apply the unchunk rule
- remove the chunk based on the given pattern.

e.g.:- S → The little cat sat on the mat.

1) Apply Pos tag

The/Dt little/Adj cat/N sat/V on/P the/Dt mat/N

2) Apply chunking → NP chunking

[The little cat] sat on [the mat]
NP NP

3) Unchunk – pattern is (DT Adj V)

⇓

The little cat

(DT Adj NN)

NP chunking

- base NPs
- used to satisfy the specialized words (nouns) in medical texts/ manuals.
- Date/Time NPs | Name NPs to identify the specialized NPs (Date/Time NPs)

e.g.:- I will meet you on [Thursday, the eighth of November at 1:30pm]

Date/Time NP.

Types of shallow Parsers

1) Chunk parser
2) FST parser
3) Machine learning parser.

1) Chunk parser

- 3 steps

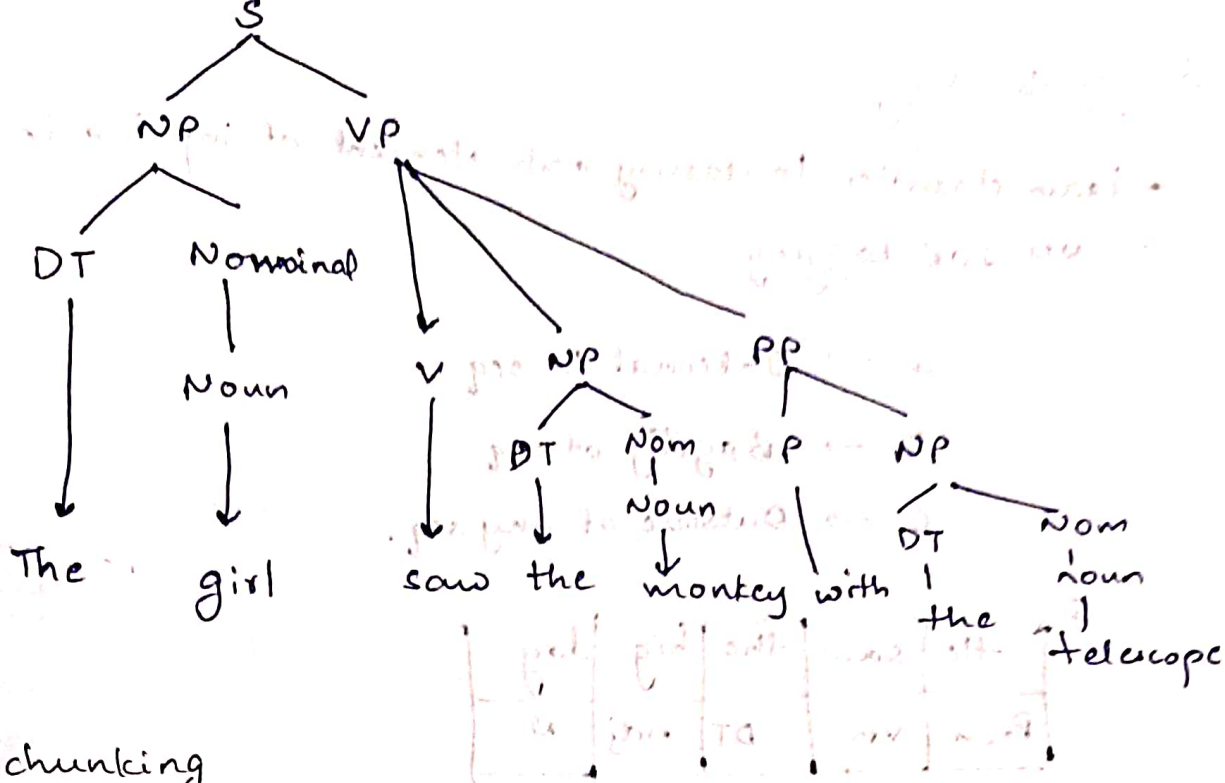
1) word identification (via PoS tagging)
2) Chunk identification (via rules)
3) ~~Attachment~~ Attachment of chunks (via parsing, grammar rules)

e.g.:- The girl saw the monkey with the telescope.

DT N VB DT N P DT N

1. (PoS tagging) → DT N VB DT N P DT N
2. (NP chunking) → [The girl] saw [the monkey] with [the telescope]

NP NP NP

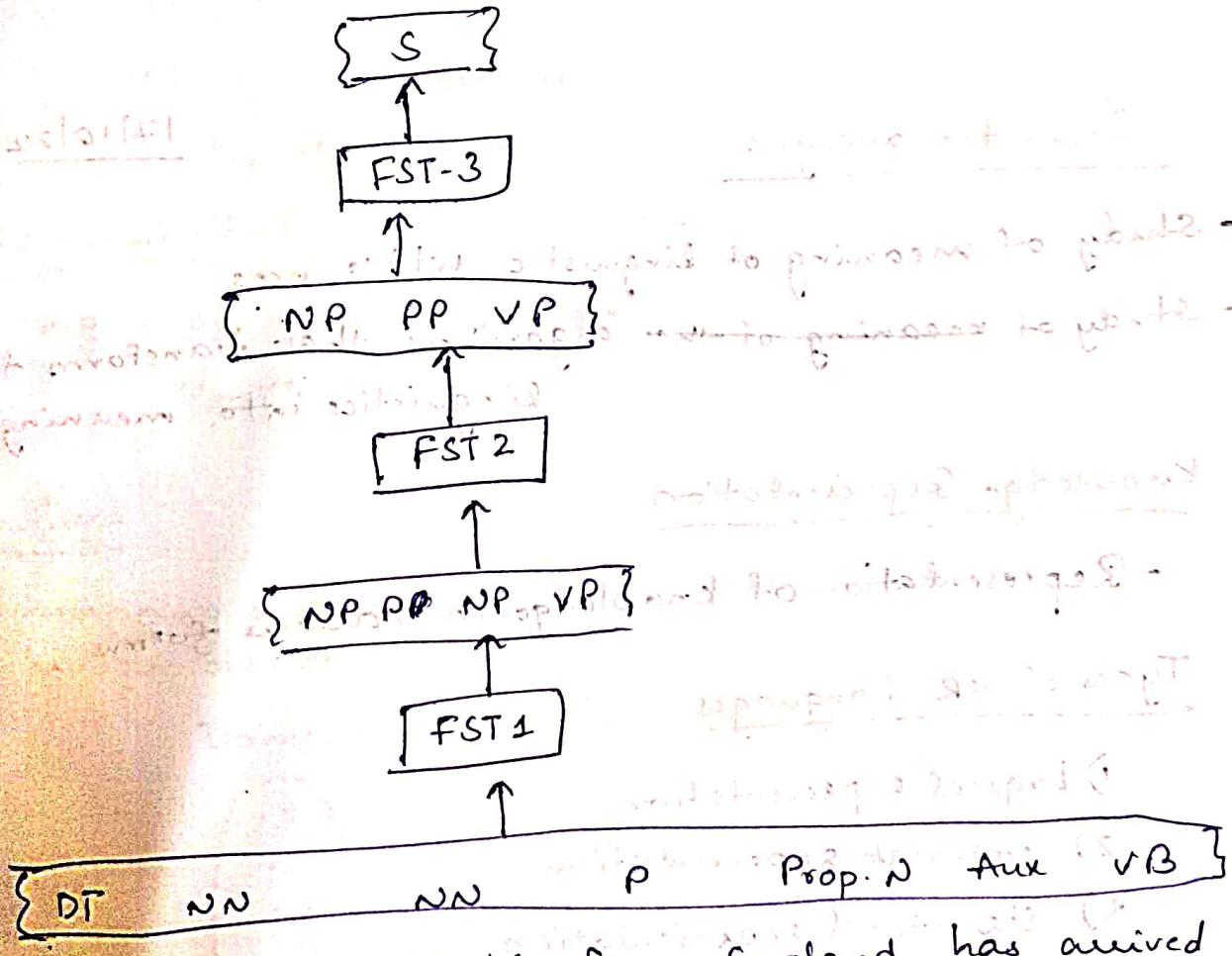


2) FST chunking

- Using Regular expressions to identify the constituents.

  e.g.:- NP → (DT) NN* NN

  - Find the longest matching chunks



3. ML chunking

- train classifier to classify each element of input x in sequence.
- uses IOB tagging

I → Internal to seq

B → Begining of seq

O → Outside of any seq.

S

He	saw	the	big	dog
Pron	VB	DT	Adj	N

B O B I I

NP NP

S

17/10/22

Semantic analysis

- Study of meaning of linguistic utterances
- Study of ~~meaning of the~~ algorithms that transform the linguistics into meaningful text

Knowledge Representation

- Representation of knowledge in various forms.

Types of KR languages

1) Logical representation
2) Network representation
3) Structural representation

1) Logical Representation

- uses mathematical knowledge to represent the knowledge.

e.g :- First order predicate calculus.

Predicate calculus

e.g :- It rained on wednesday.

weather (rain, wednesday)

Predicate :- that relates the climate - rain with the day - wednesday.

Propositional calculus

P : It rained on wednesday

↳ True / False

Quantifiers

1) Universal ∀ (for all)

e.g :- All veg restaurants serve veg dishes.

$\forall_x$ veg restaurant (x) ∧ serve (x, veg dishes)

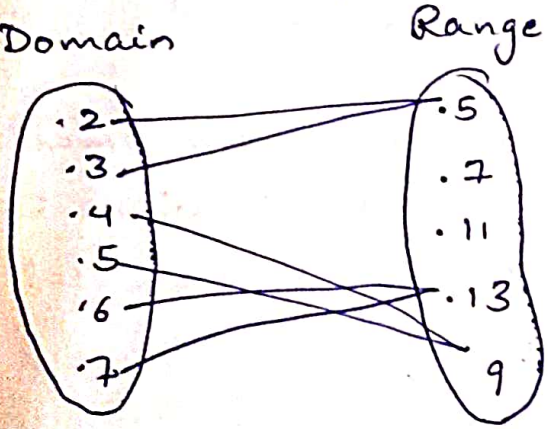
2) Existential ∃ (if there exists)

e.g :- Some people like Apples.

$\forall_x$ person(x) ∧ likes (x, Apple)

functions

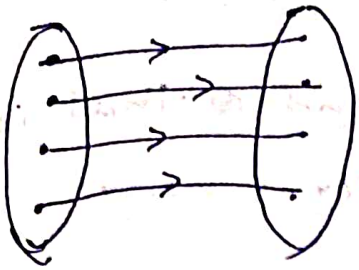
- mapping of one or more elements in domain to a unique element in range.



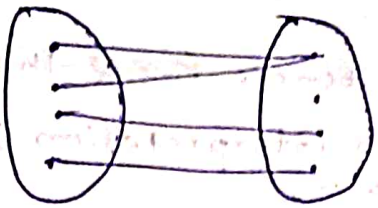
plus (2, 3) : 5

plus (4, 5) : 9

plus (6, 7) : 13



One to one mapping



many to one mapping

Connectives

→ ∧ , ∨ , ¬ , ⇒ , ⟺
(multiplication)
And OR negation equivalence

P	Q	P∧Q	P∨Q	P ⇒ Q
T	T	T	T	T
T	F	F	T	F
F	T	F	T	T
F	F	F	F	T

Inference rules

- deriving new rules from the existing ones.

1) Modus Ponxens

Consider 2 sentences P & P→Q that are true under an interpretation I, then infer a new sentence Q that is true under the same interpretation -I.

P: It is raining. (T)

P→Q: If it is raining then the ground is wet (T)

Q: The ground is wet

---

2) Modus Tollens

Consider 2 sentences $P \rightarrow Q$ that is true and Q that is false under an interpretation I, then infer a new sentence $\neg P$

e.g.:- $P \rightarrow Q$: If it is raining then the ground is wet. (T)

Q : The ground is wet (F) [The ground is dry]

$\neg P$: It is not raining

---

(2) Network representation

- uses graphs to represent the knowledge
  1) Conceptual graphs
  2) Semantic networks

Conceptual graphs

- connected, bipartite graphs
- 2 types of nodes
  1) Concept nodes → concrete, abstract
     - rep by [ ]
  2) Conceptual relational nodes
     - rep the relation b/w one or more concepts
     - → rep by ( )

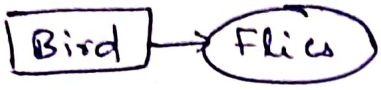
Concrete graphics concepts

- characterized by our ability to make images of them in our mind

  e.g.:- chair, Restaurant, Book

Abstract

- can't imagine their images
  - e.g.:- Punctuality, loyalty
- no labelled arcs

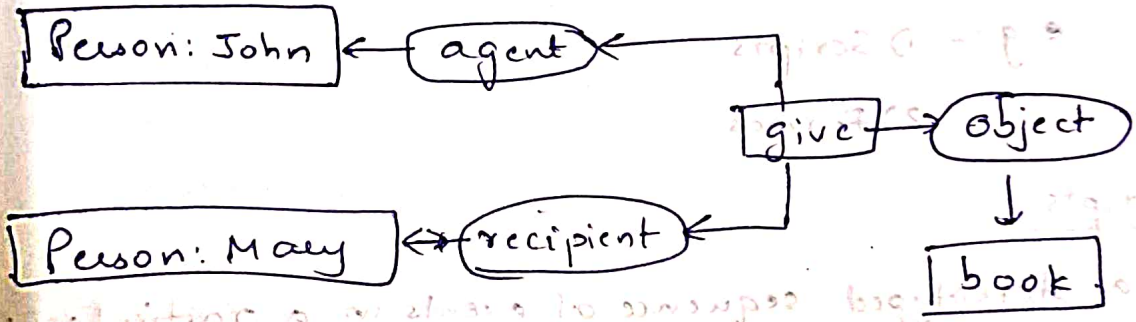
e.g.:- 

1-ary relation

2-ary

3-ary

- e.g.:- John gave Mary the book.



Semantic networks

18/10/22

- Semantic networks is a network representation
- labelled, directed graph

nodes :-

represent the objects / concepts / situations.

links :-

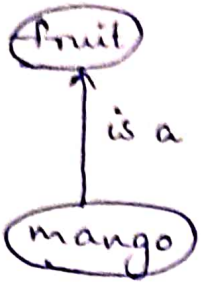
represent the relationship between concepts that contain the structural information of the knowledge to be

represented

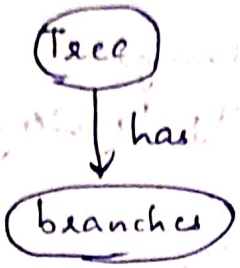
Labels – indicate the types of relationships

e.g.:-

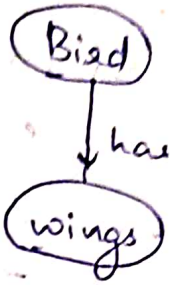
Mango is a fruit



Tree has branches



Bird has wings



3. Structural representation

- extends the network by having each node - a complex data structure having named slots with attached values.
- values → simple numeric value or a pointer to other frame or a procedure to perform a particular task.

e.g.:- 1) Scripts

2) Frames.

Scripts

- Is a stereotyped sequence of events in a particular context.

components:-

1) Entry condition – descriptions that must be true for a script to be called.

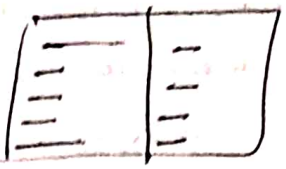
2) Results – facts that are true once a script has been terminated.

3) Props – things that support the context of the script.

4) Roles – actions performed by individual participants.

5) Scenes – is a sequence ~~of that~~ represents the temporal

aspect of the script.

Script for a restaurant system or 

SCRIPT: RESTAURANT

Props:- Tables, chairs, spoons, menu cards, food, ...

Roles:-

S: Customer
O: Owner
M: Cashier
W: Waiter
C: Cook

Entry conditions:-

S is hungry
S has money
Restaurant is open

Results:-

- S eats food
- S money gets deducted
- O earns money

SCENE1:- ENTERING

1. _____
2. _____
3. _____

SCENE2:- ORDERING

1. _____
2. _____
3. _____

SCENE3:- EATING

1. _____
2. _____
3. _____

SCENE4:- EXITING

1. _____
2. _____

Q. ~~Convert the following state~~

Write the scripts for

1) Online ~~transaction~~ train ticket reservation system.
2) Library system
3) University

Q. Convert the following sentences into first order predicate logic form.

1) All employees earning 2 lakhs or more per year must pay the income tax.
2) Some kids are naughty
3) Nobody likes taxes
4) If wishes were horses, beggars would ride.
5) If it is not raining today, then John will go to market.
6) A restaurant that serves Mexican food is nearer to airport
7) When Mary's flight departed, I ate lunch.
8) If you are tall, then you can play basketball.
9) Not all vegetarians eat eggs.
10) If the sky is cloudy then it will rain.

Q. Construct the conceptual graphs and semantic networks for the following statements.

1) Parrot is a green bird having wings to fly.
2) Alligator is a species of crocodile that lives both on land and water.
3) Ram likes all kinds of vegetarian food.
4) Mango is a kind of fruit available during the months of April and May.