Natural Language Processing and Machine Translation

Fundamentals of Natural Language Processing

Abhishek Koirala

M.Sc. in Informatics and Intelligent Systems Engineering

Beginnings of NLP

- Started in 1940s after WW2
- By 1958, significant issues were discovered with natural language
- Between 1957-1970, researchers split into two divisions of nlp
  - Symbolic approach
  - Stochastic approach

Symbolic approach

- Rule based approach
- Focused on formal languages and generating syntax

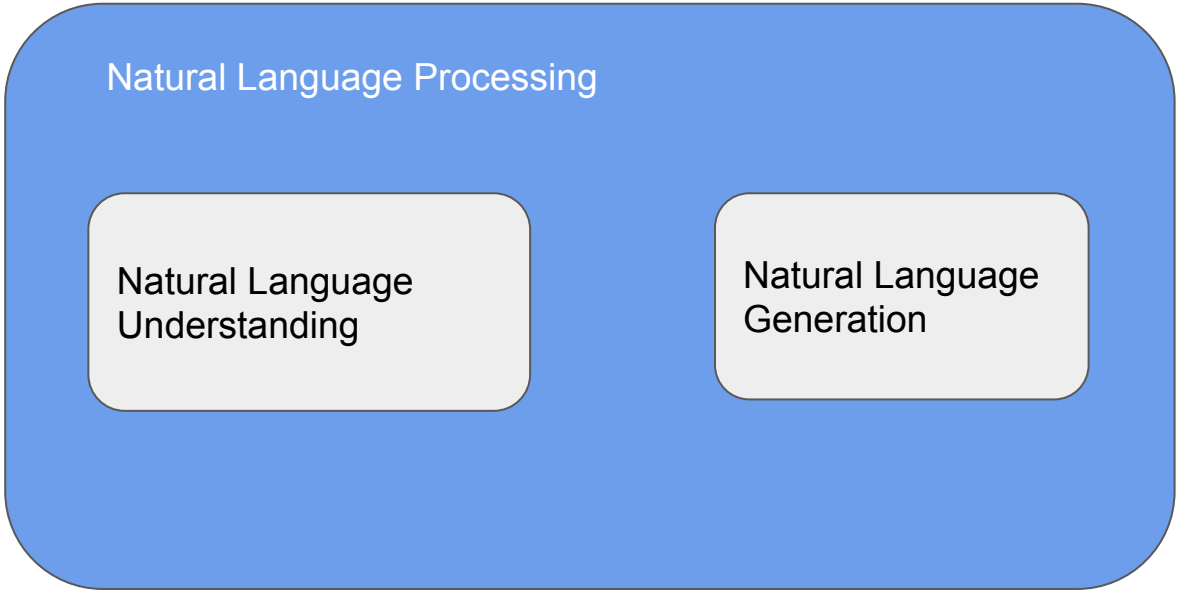
Stochastic approach

- Statistical and probabilistic methods of NLP

After 1970, **logic-based paradigms,** was embraced, that focused on encoding rules and language in mathematical logic, contributing to development of **Prolog**

THAPATHALI CAMPUS
INSTITUTES OF ENGINEERING

Generic NLP systems



Natural Language Understanding

Natural Language

NLU

- Concepts
- Entities
- Keywords
- Emotion
- Relation
- Semantic roles

THAPATHALI CAMPUS
INSTITUTES OF ENGINEERING

Phases of NLP

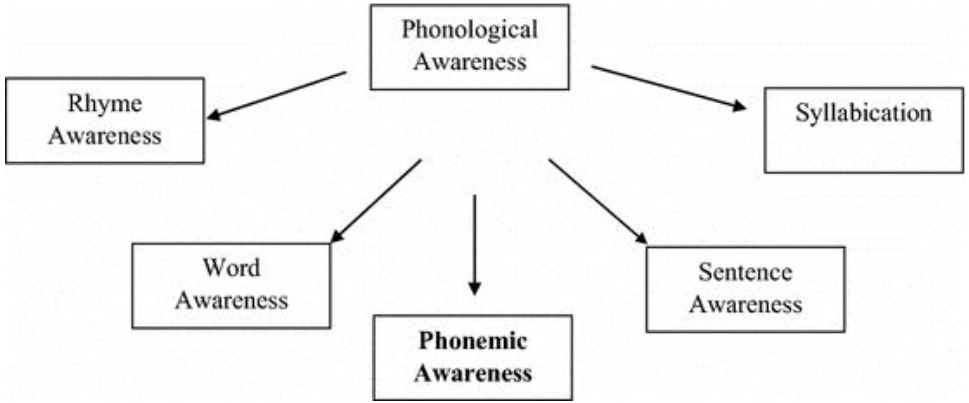
- Phonological Analysis
- Morphological Analysis
- Lexical Analysis
- Syntactic Analysis
- Semantic Analysis
- Discourse Integration
- Pragmatic Analysis

Phonological Analysis

- Study of speech and sounds of a particular language
- Very first component of language

Phonetics

- Study of how speech sounds of human language from the perspective of their production, perception or their physical properties



Morphological Analysis

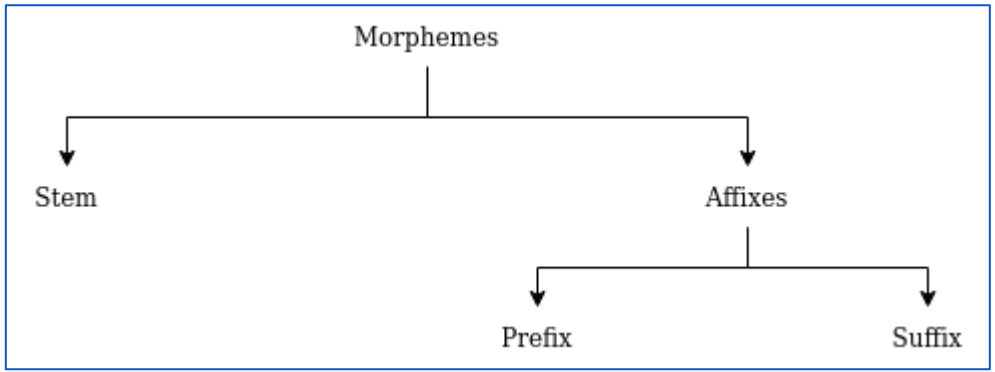
- Study of word formation
- Identification, analysis of root words, affixes, POS

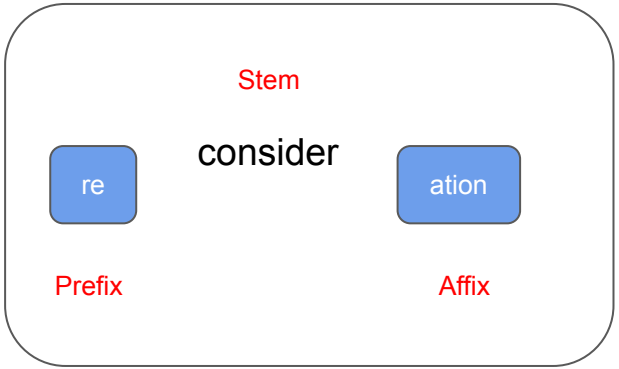
Washing = Wash + ing

Browser = Browse + er

Rats = Rat + s

Morphological Analysis





Morphological Analysis

- Recognized root form of inflected words and construct a standardized representation

book = book + SG

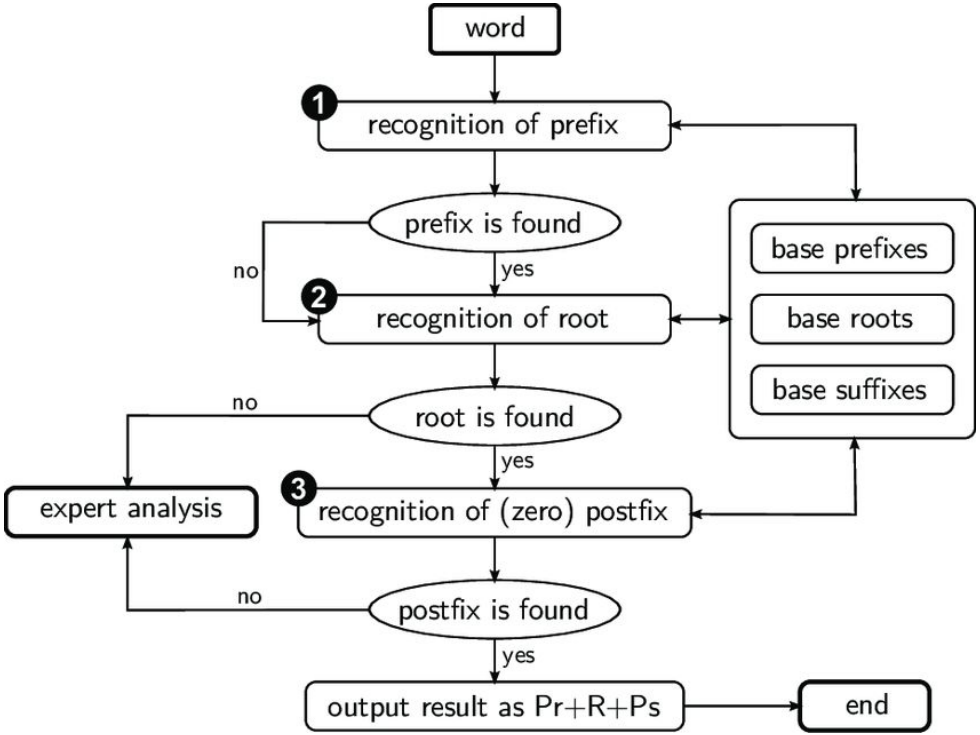
books = book + PL

ran = run + PAST

Input	Morphological
Cats	Cat + N + PL
Geese	Goose + N + PL
caught	Catch + V + PAST
missed	Miss + V + PAST
cat	cat + N + SG

THAPATHALI CAMPUS
INSTITUTES OF ENGINEERING

Morphological Analysis

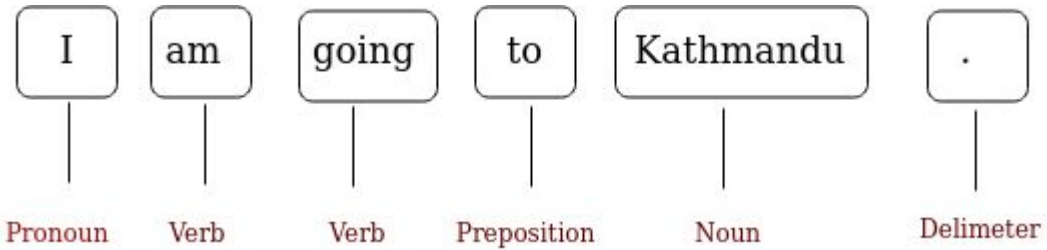


Kovár, V., Horák, A., & Jakubícek, M. (2010, December). How to Analyze Natural Language with Transparent Intensional Logic?. In *RASLAN* (pp. 69-76).

THAPATHALI CAMPUS
INSTITUTES OF ENGINEERING

Lexical Analysis

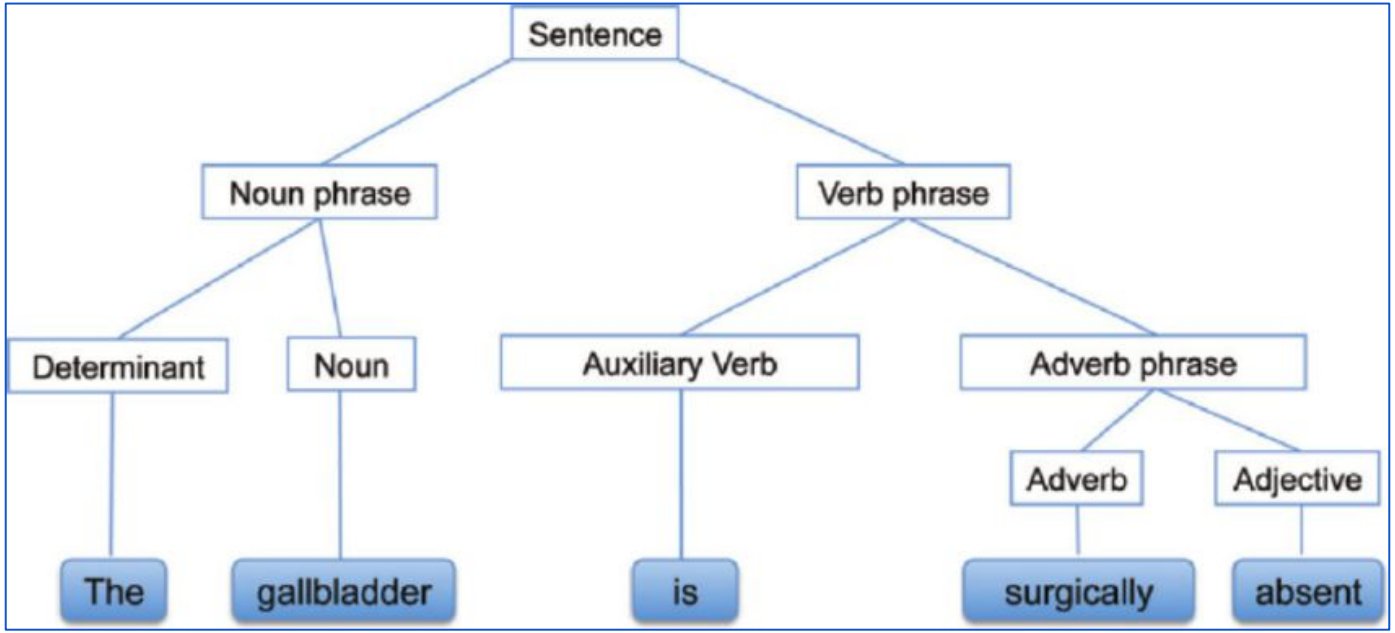
- Understand what the word means
- Understand the context
- Note the relationship of one words to others
- Lexeme vs Tokens



Lexeme	Tokens
I	PRONOUN
am	VERB
going	VERB
to	PROPOSITION
Kathmandu	NOUN
.	DELIMITER

Syntactic Analysis

- Check syntactic structure and its components
- Generates a parse tree



Cai, T., Giannopoulos, A. A., Yu, S., Kelil, T., Ripley, B., Kumamaru, K. K., ... & Mitsouras, D. (2016). Natural language processing technologies in radiology research and clinical applications. *Radiographics*, *36*(1), 176.

Syntactic Analysis

Context Free Grammar

Consists of a finite set of grammar rules represented as a quadruple (N, T, P, S) where

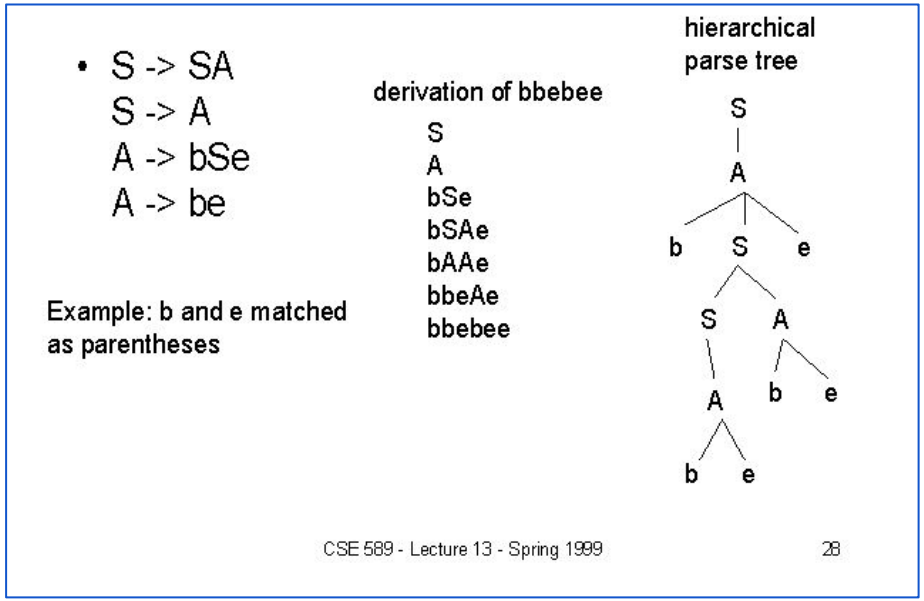
N = set of non terminal symbols

T = set of terminal symbols

P = set of rules, P : N -> (N U T)*

S = start symbol

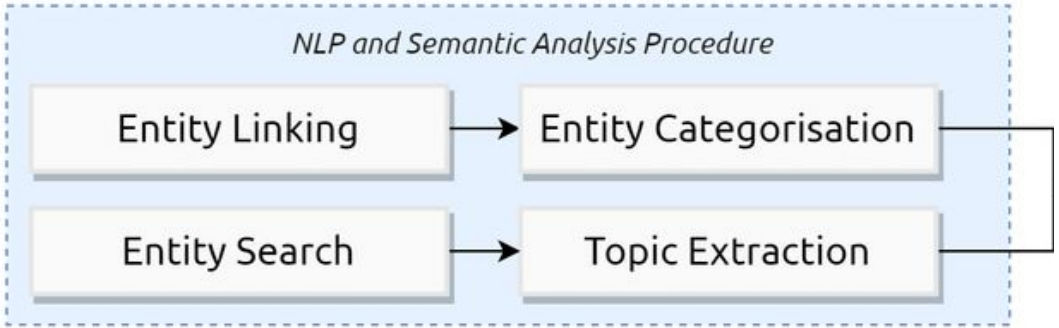
Context Free Grammar



https://courses.cs.washington.edu/courses/csep521/99sp/lectures/lecture13/sld028.htm

Semantic Analysis

- Helps in understanding the context
- Understanding the emotions that might be depicted in the sentence
- Extracting information from text
- Commonly used in translation, chatbots and search engines



López-Martínez, A., Carrera, Á., & Iglesias, C. A. (2020). Empowering museum experiences applying gamification techniques based on linked data and smart objects. *Applied Sciences*, *10*(16), 5419.

Semantic Analysis

Parts of semantic analysis

- Lexical Semantic Analysis
- Compositional Semantic Analysis

Tasks involved in Semantic analysis

- Word Sense Disambiguation
- Relationship Extraction

THAPATHALI CAMPUS
INSTITUTES OF ENGINEERING

Semantic Analysis

Elements of Semantic Analysis

- Hyponymy

  Color -> grey, blue, white, black

- Homonymy

  Rose, bank

- Synonomy

  (large, big) (tall, gigantic)

- Antonomy

  (hot, cold), (tall, short)

THAPATHALI CAMPUS
INSTITUTES OF ENGINEERING

Semantic Analysis

Elements of Semantic Analysis

- Polysemy

  Man -> “human species” , “male human” , “adult male”

- Meronomy

  Wheel -> Automobile

Semantic Analysis

Basic building blocks of semantic system

- Entities
- Concept
- Relations
- Predicates
- Case grammar

Semantic Analysis

Approaches of Semantic analysis

- **First order predicate logic**

  I love pizza. -> love(I, pizza)

  Jack studies english and mathematics -> studies(Jack, english) ^ studies(Jack, mathematics)

  Jill takes biology or physics -> takes(Jill, biology) v takes(Jill, physics)

Semantic Analysis

- **First order predicate logic**

- All students taking AI are genius.

  $\forall x$: students(x) ^ taking AI(x) → genius(x)

- Some students are taking AL are genius and hardworking.

  $\exists x$: students(x) ^ taking AI(x) →genius(x) ^hardworking(x)

- **Hiba takes DM and AI.**

  $\exists x$: hiba(x) ^ takes(x, DM) ^ takes(x, AI)

https://www.slideshare.net/HabibaSaeed5/first-order-predicate-logic-examples-different-examples

THAPATHALI CAMPUS
INSTITUTES OF ENGINEERING

Semantic Analysis

- **First order predicate logic**

- Brothers are siblings.

  $\forall x \forall y$: brother(x) → Siblings(x, y)

- Nida has no sister.

  $\sim\exists x$: Sister.Of(x, Nida)

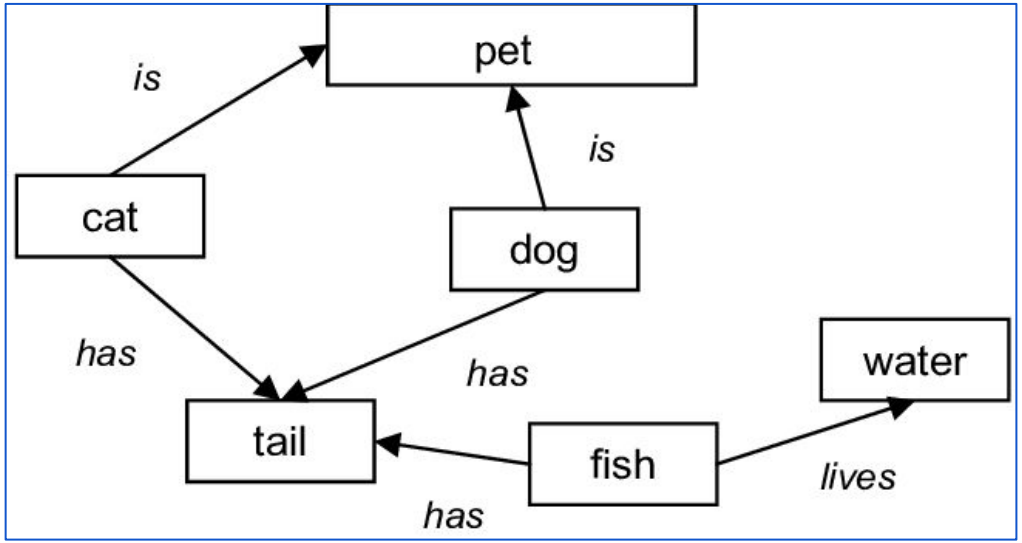
- Every student smiles.

  $\forall x$: student(x) → smile(x)

https://www.slideshare.net/HabibaSaeed5/first-order-predicate-logic-examples-different-examples

Approaches of Semantic analysis

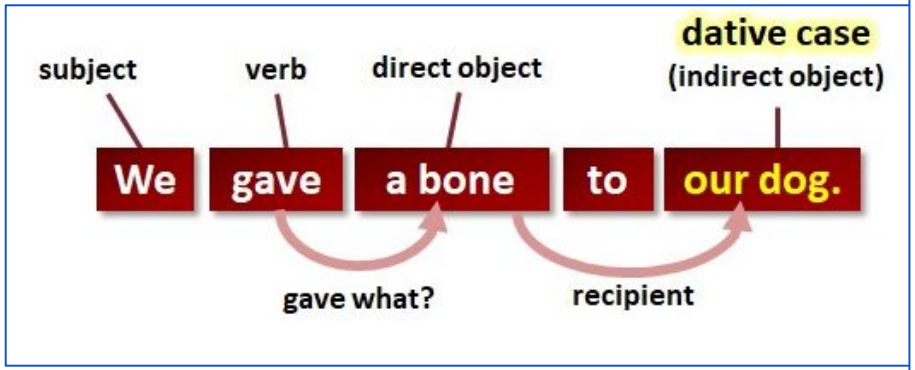
- **Semantic nets**

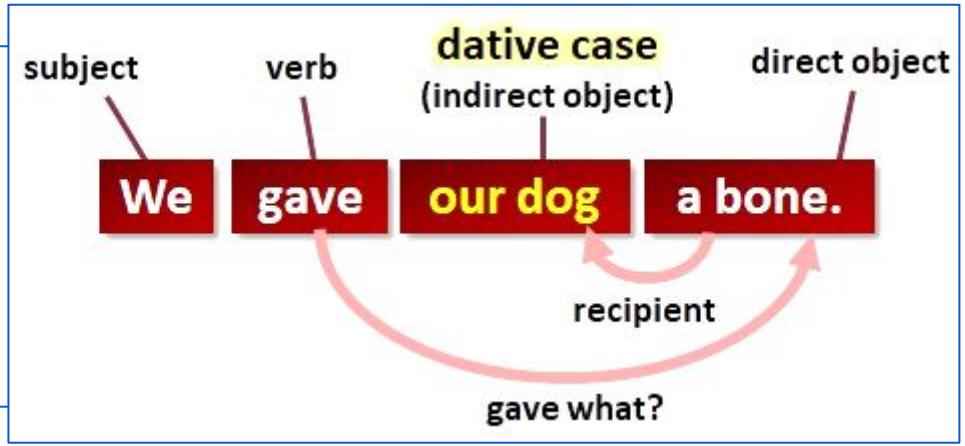


Lula, P., Morajda, J., Paliwoda-Pękosz, G., Stal, J., Tadeusiewicz, R., & Wilusz, W. (2014). Computer Methods of Data Analysis and Processing. *Computer Methods of Data Analysis and Processing, Cracow University of Economics*.

Approaches of Semantic analysis

- **Case Grammar**





https://www.grammar-monster.com/glossary/dative_case.htm

Discourse Integration

- Makes a sense of the context
- Deals with how the immediately preceding sentence can affect the interpretation of next
- who/when/where/what .... are involved in an event

Bill had a red balloon.
John wanted it.

What is **it** in this context ?

Pragmatic Analysis

- Study of language that is not directly spoken, instead hinted

Turn the heat on. It's getting cold

What does the user mean by "turning the heat on"?

Literally turning on the heat

Turning on the heater

Pragmatic Analysis

Aspects of Pragmatics

- **Deixis**

  Word or phrases that cannot be fully understood without additional information.

  Eg. Meet me here. (Meet where exactly?)

- **Implicature**

  - Conversational Implicature

    Agent 1: I am hungry
    Agent 2: There is a restaurant nearby

THAPATHALI CAMPUS
INSTITUTES OF ENGINEERING

Pragmatic Analysis

Aspects of Pragmatics

- **Presupposition**

  Based on speaker's assumption prior to making an utterance.

  Eg. John' brother is buying two mercedez

- **Speech text**

  Eg. I smashed the potato

THAPATHALI CAMPUS
INSTITUTES OF ENGINEERING

Ambiguities in NLP

- **Morphological ambiguity**
  - Word category ambiguity

    book -> Noun, book -> Verb

    bank -> Noun, bank -> Verb

- **Syntactic Ambiguity**

  Ambiguity in grammar structure.

  Eg. Visiting relatives can cause problems.
  can be interpreted as

  1) Relatives who visit can cause problems
  2) When we visit relatives, there can be problems

THAPATHALI CAMPUS
INSTITUTES OF ENGINEERING

Ambiguities in NLP

- **Semantic Ambiguity**
  - Word sense disambiguation

    Each word in a sentence has more than one meaning.

    Eg. Kanchan is currently in a terrible state

- **Discourse Ambiguity**
  - Anaphoric ambiguity

1) Monkey eats banana, when they wake up
   Who is they here? -> Monkey

2) Monkey eats banana, when they are ripe
   Who is they here? -> Banana

THAPATHALI CAMPUS
INSTITUTES OF ENGINEERING

Challenges in NLP

- Elongated words
- Shortcuts
- Handling emojis
- Mix use of language
- Ellipsis
- Punctuational ambiguity

Programming languages for NLP

- **Python**
- **Java**
- **R**

THAPATHALI CAMPUS
INSTITUTES OF ENGINEERING

Applications of NLP

- Sentiment analysis
- Text Classification
- Chatbots and virtual assistants
- Text Extraction
- Machine Translation
- Text Summarization
- Market Intelligence
- Auto-correct / Suggestions
- Intent Classification
- Speech recognition

THAPATHALI CAMPUS
INSTITUTES OF ENGINEERING


thank you
danke
謝謝
ngiyabonga
شكراً جزيلاً
рахмат
Баярлалаа
спасибо
faafetai lava
mersi
kia ora
barka
welalin
tack
spas
teşekkür ederim
mahalo
tapadh leat
vinaka
спасибі
blagodaram
dank je
misaotra
matondo
paldies
grazzi
gracias
хвала
kiitos
dankie
dhanyavad
nanni
nandri
maururu
köszönöm
hvala
gràcie
bayarlalaa
bedankt
enkosi
dziękuję
asante
manana
obrigada
tenki
murakoze
djiere dieuf
tau
mochchakkeram
дякую
mamnun
chokrane
aciu
dankon
akun
chnorakaloutioun
gratias ago
gràcies
sulpáy
taiku
go raibh maith agat
sobodi
děkuji
obrigado
mèsi
didi madloba
sagolun
sukriya
kop khun krap
ありがとう
najis tuke
kam sah hamnida
rahmat
terima kasih
tanemirt
rahmet
grazie
arigatô
takk
dakujem
trugarez
diolch
dhanyavadagalu
shukriya
merce
мерси
তোমাকে ধন্যবাদ
감사합니다
xièxie
ευχαριστώ
merci