Introduction to
Natural Language Processing

Tanmoy Chakraborty
Associate Professor, IIT Delhi
https://tanmoychak.com/

Introduction to Large Language Models

THE TIMES OF INDIA

INCLUSIVE OF EDUCATION TIMES & DELHI TIMES (CIRCULATED ONLY IN DELHI NCR) | *APPLICABLE ONLY ON MONTHLY PURCHASE (IN DELHI NCR)

INDIA'S LARGEST ENGLISH NEWSPAPER

PM Modi crosses 100m followers on X, cementing his position as the most followed active politician globally on the social networking site. His handle has witnessed growth of 30m followers in 3 years, P 10

Sanju Samson smashes his second T20I fifty, while fast bowler Mukesh Kumar picks his career-best figures of 4/22 in the format as **India beat Zimbabwe by 42 runs in the final game to win series 4-1,** P17

Indian cricket board to release **₹1cr in financial assistance to ex-cricketer and coach Anshuman Gaekwad,** who has been battling blood cancer for over a year, P 17

CPN-UML chairman **K P Sharma Oli, who's seen as pro-China, appointed Nepal's PM for a fourth term** to lead the new coalition govt that faces the daunting challenge of providing political stability in the Himalayan nation, P 16

IN THE COURTS

- **Madras HC quashes punishment imposed on a constable for sporting a beard** in accordance with his religious beliefs, ruling that disciplinary action was 'shockingly disproportionate' | P 8
- **Bombay HC directs RPO to reissue a Mumbai resident a passport** after it rejected his application on a **wrong police verification report** | P 10
- **A Delhi court acquits a man accused of rape,** stating that the victim's testimonies are neither clear, cogent, credible, nor trustworthy | P 7

Donald Trumps Death

■ Bullet Pierces Ear At Rally ■ Secret Service Kills Shooter ■ 1 Rallygoer Dead, 2 Injured

Chidanand Rajghatta | TNN

Washington: Donald Trump, US ex-president and Republican candidate for this year's presidential election, escaped by centimetres an assassination attempt during a rally in Pennsylvania on Saturday, convulsing an already turbulent political scene in America.

Bullets fired by a lone gunman positioned on a nearby rooftop nicked Trump's right ear and bloodied it — he was later said to be "fine and in great spirits" — but a 50-year-old man, besides the assailant who was immediately shot dead by the Secret Service, was killed in the incident. Two other rallygoers were critically injured.

In iconic images immediately flashed across the world, a fearless Trump, breaking free from a huddle of Secret Service agents protecting him, raised a clenched fist with blood streaking across his cheek. The indelible moment, consecrated into campaign merchandise within hours of the incident, inflamed and galvanised Trump's supporters, and is expected to power him

▶EDIT PAGE: **Trump's Moment/Shot That'll Divide America More**

DOWN & BACK UP: 'FIGHT', SAYS DEFIANT DON

Getty Images/AP

Trump drops to the ground (L) after being shot at. He then stands up and pumps his fist, making for an instantly iconic picture

- At 6.02pm Saturday, Donald Trump takes the stage. Soon, **2 spectators spot an armed man atop a building, raise alarm**
- **Shots ring out** at 6.08pm. **Trump grabs his right ear with his hand.** More shots heard
- Secret Service agents cover him, shout: 'Get down'. **Trump crouches** behind the lectern
- 17 seconds after the 1st shot, final pop is heard and a woman screams. 8 shots fired in all. **Agents neutralise shooter**
- Trump stands up, **face streaked with blood.** Agents try ushering him offstage. 'Wait, wait, wait,' he tells them
- Trump pumps his fist and says 'Fight! Fight!' Crowd chants, 'USA, USA'. Agents hustle him into SUV, drive away
- The ex-president later says, 'The **bullet pierced the upper part of my right ear. I felt it ripping through my skin. Much bleeding took place'**

■ **PM Modi says he's 'deeply concerned by attack on my friend' Trump, condemns it**

■ **15 direct attacks on US presidents/ ex-prez/presidential candidates** P 16

Motive of shooter, a Republican who would've been 1st-time voter in Nov prez polls, 'unclear'

The motive of shooter Thomas Matthew Crooks is still unclear,

Secret Service in line of fire for 'security failures', FBI agent says breach 'surprising'

The much-vaunted Secret Service, often portrayed in heroic

Gang guns down Trinamool man at Bengal dhaba

Bapi Ray (36), a Trinamool member, was shot dead and another seriously wounded when a gang of eight to 10 people fired at them while they were dining at a roadside dhaba at Islampur in Bengal's Uttar Dinajpur. Business rivalry could be the murder motive, sources said. The killing sparked sporadic protests on Sunday. P 8

City woman, BSES staffer electrocuted

A woman, who went to a hospital to visit her son Saturday, died of electrocution in a waterlogged street in Bhajanpura. In another incident, a BSES staffer check-

32-year-old patient shot dead 'by teen' in GTB hosp ward

Killing A Case Of Mistaken Identity: Kin

Abhay@timesofindia.com

New Delhi: A 32-year-old patient was shot dead, allegedly by an 18-year-old youth, inside a ward of GTB Hospital in Shahdara Sunday. His family has claimed he was killed in a case of mistaken identity and that the intended target was a history-sheeter, who was admitted to the same ward.

The victim, Riyazuddin, was a labourer who lived with his family in Sriram Nagar, Khajuri Khas. He had been admitted to the hospital on June 23 for treatment of an abdominal infection.

Police personnel investigate at the hospital on Sunday

Sunday's incident took place around 4pm, a senior police officer said. The suspect allegedly came to ward number 24 and fired at least two rounds at the patient, who was receiving dressing from the nurse.

▶20 people, P 3

CRPF jawan killed, 2 cops injured in Manipur attack

Kangkan.Kalita1 @timesofindia.com

Guwahati: A CRPF constable from Bihar was killed and two Manipur cops and an unidentified civilian were wounded Sunday in an ambush by

CIVILIAN WOUNDED

- Patrol team bombarded by **militants, strategically positioned at 5-6 locations**
- CRPF's **Ajay Jha at the wheel, first to be struck**

Natural Language Processing

What is a Natural Language?

Any language that has evolved naturally in humans through use and repetition without conscious planning or pre-meditation.

Content credits: https://www.javatpoint.com/nlp

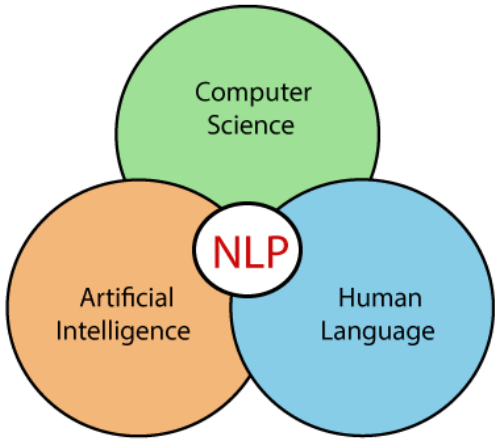
Natural Language Processing

What is a Natural Language?

Any language that has evolved naturally in humans through use and repetition without conscious planning or pre-meditation.

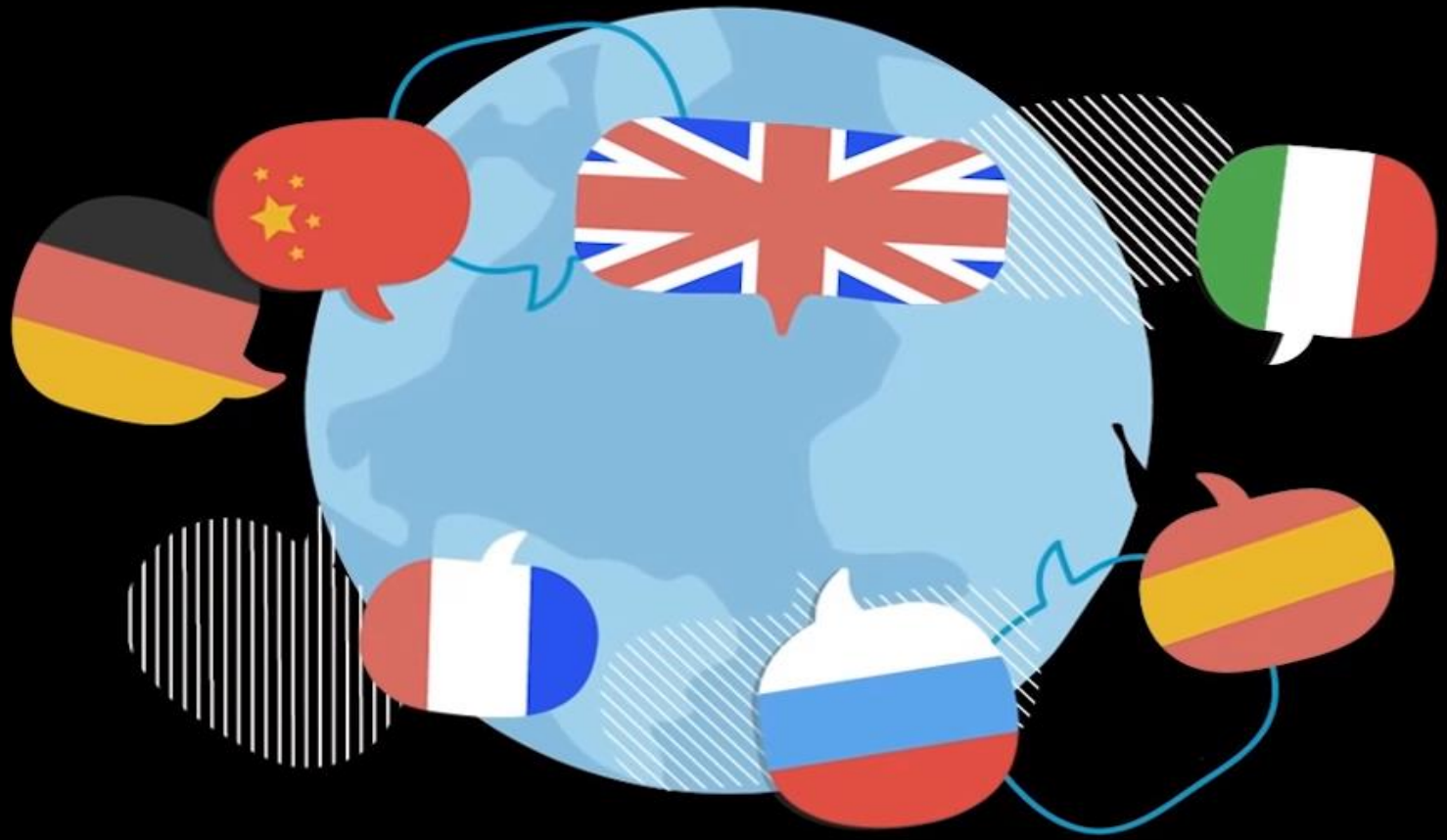
What is a Natural Language Processing?

A field of computer science, artificial intelligence and computational linguistics concerned with the interactions between computers and human (natural) languages.



Content credits: https://www.javatpoint.com/nlp

The Human Language



6500 LANGUAGES

The Human Language

Home / India / More than 19,500 mother tongues spoken in India: Census

More than 19,500 mother tongues spoken in India: Census

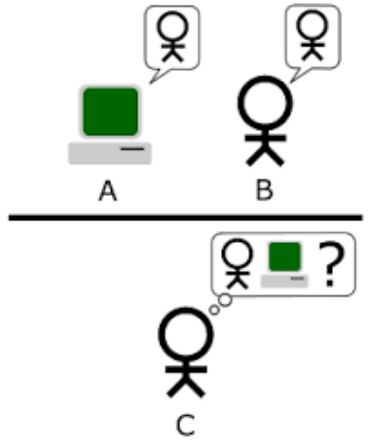
There are 121 languages which are spoken by 10,000 or more people in India, which has a population of 121 crore, the report said.

https://indianexpress.com/article/india/more-than-19500-mother-tongues-spoken-in-india-census-5241056/

Natural Language Processing

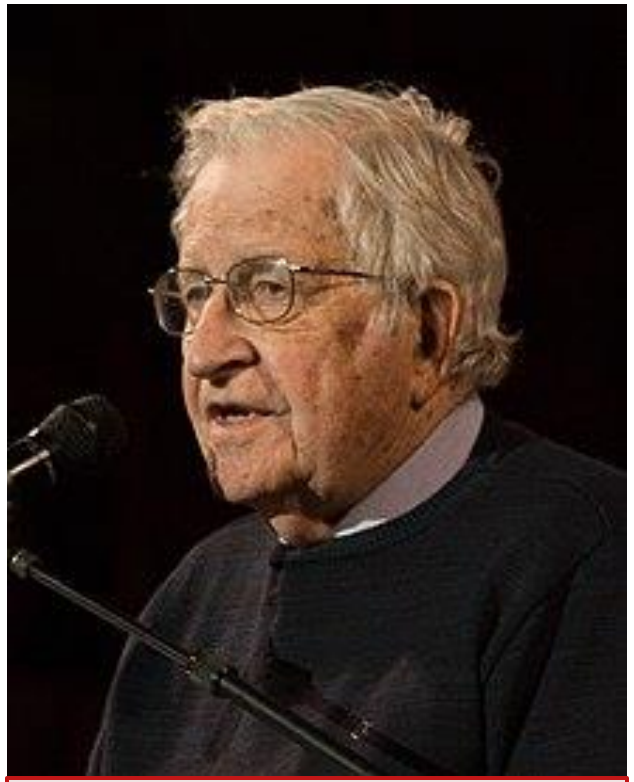
Setup

- Two rooms, two humans, and a computer.
  - Room 1: One human C
  - Room 2: One computer (A) and one human (B)
- Response generated from room 2 (either by A or B)
- C has to figure out the source of the response
  - If C is successful → "A" failed the Turing test
  - Else, → "A" passed the Turing test



"Computing Machinery and Intelligence" proposed what is now called the Turing test .

Natural Language Processing



The father of modern linguistics

In 1957, **Noam Chomsky**'s **Syntactic Structures** revolutionized Linguistics with '**universal grammar**', a rule-based system of syntactic structures

He is a laureate professor of linguistics at University of Arizona and an institute professor emeritus at MIT.

Why is NLP Challenging?

Why is NLP Challenging?

Ambiguity

The Real Reason Why NLP is Hard

THE TIMES OF INDIA

INCLUSIVE OF EDUCATION TIMES & DELHI TIMES (CIRCULATED ONLY IN DELHI NCR) | *APPLICABLE ONLY ON MONTHLY PURCHASE (IN DELHI NCR)

INDIA'S LARGEST ENGLISH NEWSPAPER

PM Modi crosses 100m followers on X, cementing his position as the most followed active politician globally on the social networking site. His handle has witnessed growth of 30m followers in 3 years, P 10

Sanju Samson smashes his second T20I fifty, while fast bowler Mukesh Kumar picks his career-best figures of 4/22 in the format as India beat Zimbabwe by 42 runs in the final game to win series 4-1, P17

Indian cricket board to release ₹1cr in financial assistance to ex-cricketer and coach Anshuman Gaekwad, who has been battling blood cancer for over a year, P 17

CPN-UML chairman K P Sharma Oli, who's seen as pro-China, appointed Nepal's PM for a fourth term to lead the new coalition govt that faces the daunting challenge of providing political stability in the Himalayan nation, P 16

IN THE COURTS

- Madras HC quashes punishment imposed on a constable for sporting a beard in accordance with his religious beliefs, ruling that disciplinary action was 'shockingly disproportionate' | P 8
- Bombay HC directs RPO to reissue a Mumbai resident a passport after it rejected his application on a wrong police verification report | P 10
- A Delhi court acquits a man accused of rape, stating that the victim's testimonies are neither clear, cogent, credible, nor trustworthy | P 7

Donald Trumps Death

■ Bullet Pierces Ear At Rally ■ Secret Service Kills Shooter ■ 1 Rallygoer Dead, 2 Injured

Chidanand Rajghatta | TNN

Washington: Donald Trump, US ex-president and Republican candidate for this year's presidential election, escaped by centimetres an assassination attempt during a rally in Pennsylvania on Saturday, convulsing an already turbulent political scene in America.

Bullets fired by a lone gunman positioned on a nearby rooftop nicked Trump's right ear and bloodied it — he was later said to be "fine and in great spirits" — but a 50-year-old man, besides the assailant who was immediately shot dead by the Secret Service, was killed in the incident. Two other rallygoers were critically injured.

In iconic images immediately flashed across the world, a fearless Trump, breaking free

▶EDIT PAGE: Trump's Moment/Shot That'll Divide America More

from a huddle of Secret Service agents protecting him, raised a clenched fist with blood streaking across his cheek. The indelible moment, consecrated into campaign merchandise within hours of the incident, inflamed and galvanised Trump's supporters, and is expected to power him

DOWN & BACK UP: 'FIGHT', SAYS DEFIANT DON

Trump drops to the ground (L) after being shot at. He then stands up and pumps his fist, making for an instantly iconic picture

- At 6.02pm Saturday, Donald Trump takes the stage. Soon, 2 spectators spot an armed man atop a building, raise alarm
- Shots ring out at 6.08pm. Trump grabs his right ear with his hand. More shots heard
- Secret Service agents cover him, shout: 'Get down'. Trump crouches behind the lectern
- 17 seconds after the 1st shot, final pop is heard and a woman screams. 8 shots fired in all. Agents neutralise shooter
- Trump stands up, face streaked with blood. Agents try ushering him offstage. 'Wait, wait, wait,' he tells them
- Trump pumps his fist and says 'Fight! Fight!' Crowd chants, 'USA, USA'. Agents hustle him into SUV, drive away
- The ex-president later says, 'The bullet pierced the upper part of my right ear. I felt it ripping through my skin. Much bleeding took place'

■ PM Modi says he's 'deeply concerned by attack on my friend' Trump, condemns it

■ 15 direct attacks on US presidents/ ex-prez/presidential candidates P 16

Motive of shooter, a Republican who would've been 1st-time voter in Nov prez polls, 'unclear'

The motive of shooter Thomas Matthew Crooks is still unclear,

Secret Service in line of fire for 'security failures', FBI agent says breach 'surprising'

The much-vaunted Secret Service, often portrayed in heroic

Gang guns down Trinamool man at Bengal dhaba

Bapi Ray (36), a Trinamool member, was shot dead and another seriously wounded when a gang of eight to 10 people fired at them while they were dining at a roadside dhaba at Islampur in Bengal's Uttar Dinajpur. Business rivalry could be the murder motive, sources said. The killing sparked sporadic protests on Sunday. P 8

City woman, BSES staffer electrocuted

A woman, who went to a hospital to visit her son Saturday, died of electrocution in a waterlogged street in Bhajanpura. In another incident, a BSES staffer check-

32-year-old patient shot dead 'by teen' in GTB hosp ward

Killing A Case Of Mistaken Identity: Kin

Abhay@timesofindia.com

Police personnel investigate at the hospital on Sunday

New Delhi: A 32-year-old patient was shot dead, allegedly by an 18-year-old youth, inside a ward of GTB Hospital in Shahdara Sunday. His family has claimed he was killed in a case of mistaken identity and that the intended target was a history-sheeter, who was admitted to the same ward.

The victim, Riyazuddin, was a labourer who lived with his family in Sriram Nagar, Khajuri Khas. He had been admitted to the hospital on June 23 for treatment of an abdominal infection.

Sunday's incident took place around 4pm, a senior police officer said. The suspect allegedly came to ward number 24 and fired at least two rounds at the patient, who was receiving dressing from the nurse.

▶20 people, P 3

CRPF jawan killed, 2 cops injured in Manipur attack

Kangkan.Kalita1 @timesofindia.com

Guwahati: A CRPF constable from Bihar was killed and two Manipur cops and an unidentified civilian were wounded Sunday in an ambush by

CIVILIAN WOUNDED

- Patrol team bombarded by militants, strategically positioned at 5-6 locations
- CRPF's Ajay Jha at the wheel, first to be struck

The Real Reason Why NLP is Hard

surpass, outdo, or beat

THE TIMES OF INDIA

INCLUSIVE OF EDUCATION TIMES & DELHI TIMES (CIRCULATED ONLY IN DELHI NCR) | *APPLICABLE ONLY ON MONTHLY PURCHASE (IN DELHI NCR)

...GEST ENGLISH NEWSPAPER

PM Modi crosses 100m followers on X, cementing his position as the most followed active politician globally on the social networking site. His handle has witnessed growth of 30m followers in 3 years, P 10

Sanju Samson smashes his second T20I fifty, while fast bowler Mukesh Kumar picks his career-best figures of 4/22 in the format as India beat Zimbabwe by 42 runs in the final game to win series 4-1, P17

Indian cricket board to release ₹1cr in financial assistanc... ex-cricketer and ce... ...Gaekwad ...been battling blo... ...cer for over a year, P 17

...UML chairman K P Sharma Oli, who's seen as pro-China, appointed Nepal's PM for a fourth term to lead the new coalition govt that faces the daunting challenge of providing political stability in the Himalayan nation, P 16

IN THE COURTS

- Madras HC quashes punishment imposed on a constable for sporting a beard in accordance with his religious beliefs, ruling that disciplinary action was 'shockingly disproportionate' | P 8
- Bombay HC directs RPO to reissue a Mumbai resident a passport after it rejected his application on a wrong police verification report | P 10
- A Delhi court acquits a man accused of rape, stating that the victim's testimonies are neither clear, cogent, credible, nor trustworthy | P 7

Donald Trumps Death

■ Bullet Pierces Ear At Rally ■ Secret Service Kills Shooter ■ 1 Rallygoer Dead, 2 Injured

Chidanand Rajghatta | TNN

Washington: Donald Trump, US ex-president and Republican candidate for this year's presidential election, escaped by centimetres an assassination attempt during a rally in Pennsylvania on Saturday, convulsing an already turbulent political scene in America.

Bullets fired by a lone gunman positioned on a nearby rooftop nicked Trump's right ear and bloodied it — he was later said to be "fine and in great spirits" — but a 50-year-old man, besides the assailant who was immediately shot dead by the Secret Service, was killed in the incident. Two other rallygoers were critically injured.

In iconic images immediately flashed across the world, a fearless Trump, breaking free

▶EDIT PAGE: **Trump's Moment/Shot That'll Divide America More**

from a huddle of Secret Service agents protecting him, raised a clenched fist with blood streaking across his cheek. The indelible moment, consecrated into campaign merchandise within hours of the incident, inflamed and galvanised Trump's supporters, and is expected to power him

DOWN & BACK UP: 'FIGHT', SAYS DEFIANT DON

Getty Images/AP

Trump drops to the ground (L) after being shot at. He then stands up and pumps his fist, making for an instantly iconic picture

- At 6.02pm Saturday, Donald Trump takes the stage. Soon, **2 spectators spot an armed man atop a building, raise alarm**
- **Shots ring out** at 6.08pm. **Trump grabs his right ear with his hand.** More shots heard
- Secret Service agents cover him, shout: 'Get down'. **Trump crouches** behind the lectern
- 17 seconds after the 1st shot, final pop is heard and a woman screams. 8 shots fired in all. **Agents neutralise shooter**
- Trump stands up, **face streaked with blood.** Agents try ushering him offstage. 'Wait, wait, wait,' he tells them
- Trump pumps his fist and says 'Fight! Fight!' Crowd chants, 'USA, USA'. Agents hustle him into SUV, drive away
- The ex-president later says, 'The **bullet pierced the upper part of my right ear. I felt it ripping through my skin. Much bleeding took place'**

■ **PM Modi says he's 'deeply concerned by attack on my friend' Trump, condemns it**

■ **15 direct attacks on US presidents/ ex-prez/presidential candidates** P 16

Motive of shooter, a Republican who would've been 1st-time voter in Nov prez polls, 'unclear'

The motive of shooter Thomas Matthew Crooks is still unclear,

Secret Service in line of fire for 'security failures', FBI agent says breach 'surprising'

The much-vaunted Secret Service, often portrayed in heroic

32-year-old patient shot dead 'by teen' in GTB hosp ward

Killing A Case Of Mistaken Identity: Kin

Abhay@timesofindia.com

New Delhi: A 32-year-old patient was shot dead, allegedly by an 18-year-old youth, inside a ward of GTB Hospital in Shahdara Sunday. His family has claimed he was killed in a case of mistaken identity and that the intended target was a history-sheeter, who was admitted to the same ward.

The victim, Riyazuddin, was a labourer who lived with his family in Sriram Nagar, Khajuri Khas. He had been admitted to the hospital on June 23 for treatment of an abdominal infection.

Police personnel investigate at the hospital on Sunday

Sunday's incident took place around 4pm, a senior police officer said. The suspect allegedly came to ward number 24 and fired at least two rounds at the patient, who was receiving dressing from the nurse.

▶20 people, P 3

CRPF jawan killed, 2 cops injured in Manipur attack

Kangkan.Kalita1 @timesofindia.com

Guwahati: A CRPF constable from Bihar was killed and two Manipur cops and an unidentified civilian were wounded Sunday in an ambush by

CIVILIAN WOUNDED

- Patrol team bombarded by **militants, strategically positioned at 5-6 locations**
- CRPF's **Ajay Jha at the wheel, first to be struck**

Gang guns down Trinamool man at Bengal dhaba

Bapi Ray (36), a Trinamool member, was shot dead and another seriously wounded when a gang of eight to 10 people fired at them while they were dining at a roadside dhaba at Islampur in Bengal's Uttar Dinajpur. Business rivalry could be the murder motive, sources said. The killing sparked sporadic protests on Sunday. P 8

City woman, BSES staffer electrocuted

A woman, who went to a hospital to visit her son Saturday, died of electrocution in a waterlogged street in Bhajanpura. In another incident, a BSES staffer check-

The Real Reason Why NLP is Hard

Virat Kohli was on fire last night. He totally destroyed the other team.

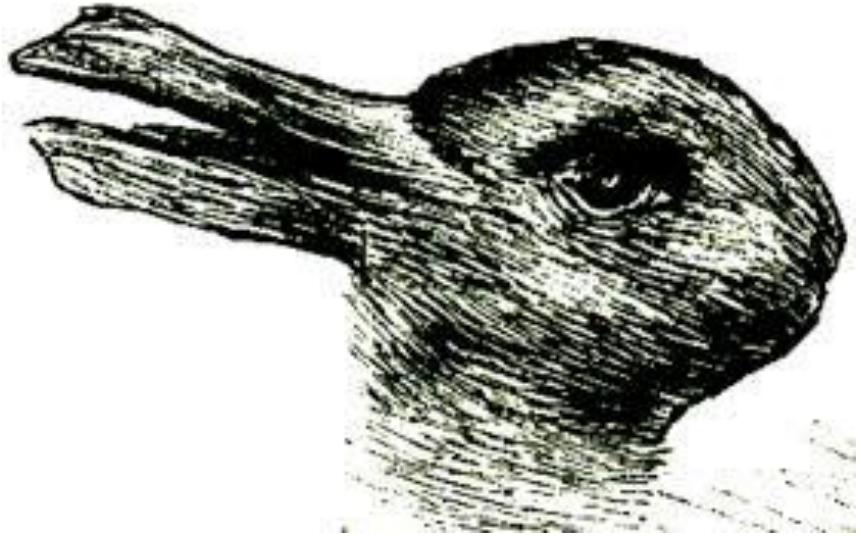
Ambiguity

Is ambiguity present in language only?

Ambiguity

Is ambiguity present in language only?

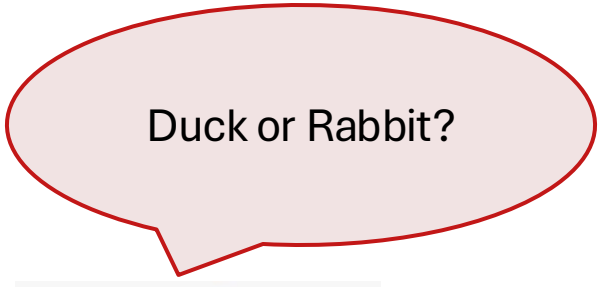
No, ambiguity is prevalent in every dimension!

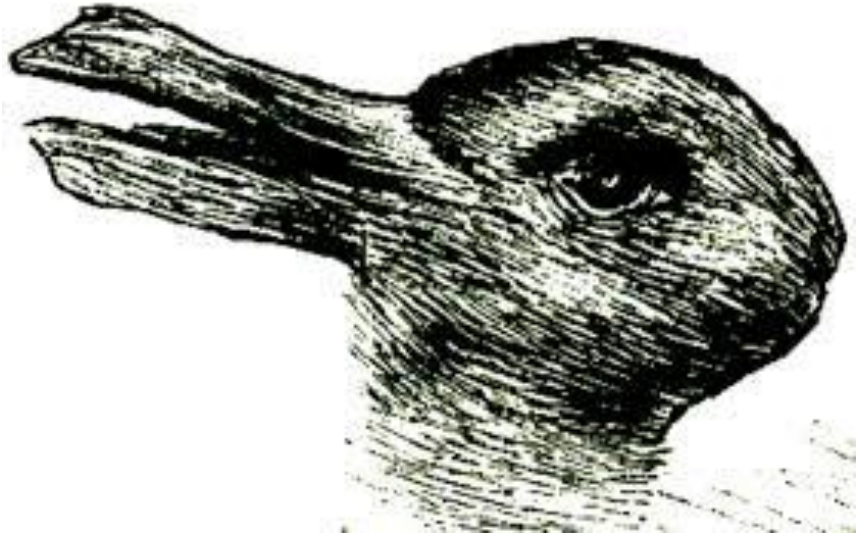


Ambiguity

Is ambiguity present in language only?

No, ambiguity is prevalent in every dimension!



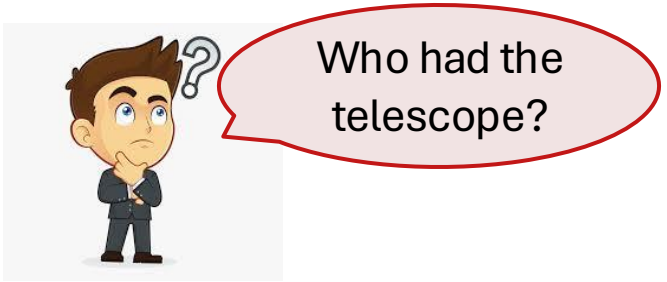


Ambiguity in Language

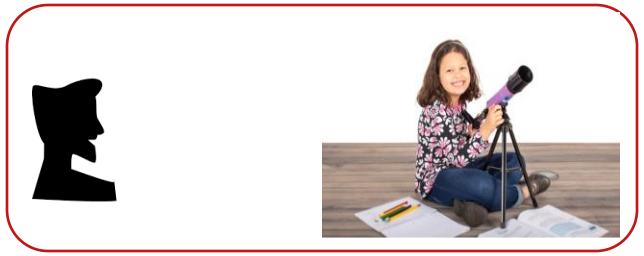
- I saw a girl with a telescope.

Ambiguity in Language

- I saw a girl with a telescope.



OR

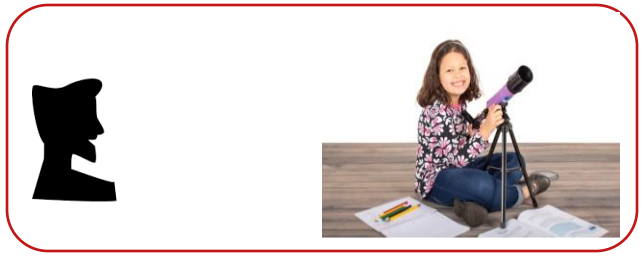


Ambiguity in Language

- I saw a girl with a telescope.

Who had the telescope?

OR



- I saw a girl with a bicycle.

- I saw a bus with a telescope.

No ambiguity!

Ambiguity in Language

- I saw a girl with a telescope.
- Mary had a little lamb.

Ambiguity in Language

- I saw a girl with a telescope.
- Mary had a little lamb.

Ambiguity in Language

- I saw a girl with a telescope.
- Mary had a little lamb.
  - Mary was physically bringing a lamb to a location, such as a farm or a home

Ambiguity in Language

- I saw a girl with a telescope.
- Mary had a little lamb.
  - Mary was physically bringing a lamb to a location, such as a farm or a home
  - Mary ate a lamb.

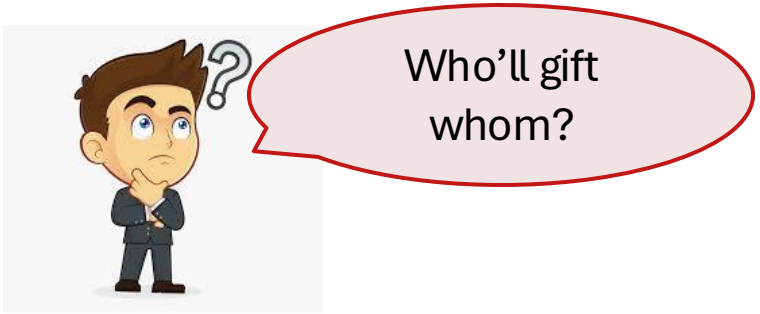
OR

Ambiguity in Language

- I saw a girl with a telescope.
- Mary had a little lamb.
- Mujhe aapko mithai khilani padegi.

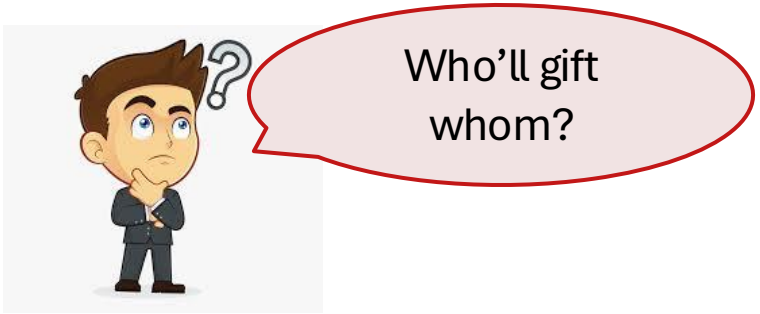
Ambiguity in Language

- I saw a girl with a telescope.
- Mary had a little lamb.
- Mujhe aapko mithai khilani padegi.



Ambiguity in Language

- I saw a girl with a telescope.
- Mary had a little lamb.
- Mujhe aapko mithai khilani padegi.



I have to gift you some sweets.

OR

You have to gift me some sweets.

Ambiguity in Language

- I saw a girl with a telescope.
- Mary had a little lamb.
- Mujhe aapko mithai khilani padegi.
- I ate rice with spoon.
- I ate rice with curd.
- I ate rice with Rahul.

Ambiguity in Language

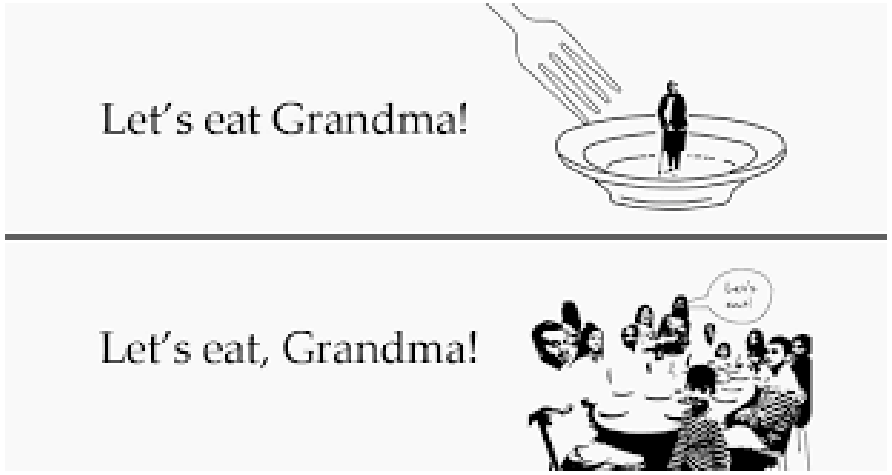
- I saw a girl with a telescope.
- Mary had a little lamb.
- Mujhe aapko mithai khilani padegi.
- I ate rice with spoon.
- I ate rice with curd.
- I ate rice with Rahul.

Ambiguity in Language

- I saw a girl with a telescope.
- Mary had a little lamb.
- Mujhe aapko mithai khilani padegi.
- I ate rice with spoon.
- I ate rice with curd.
- I ate rice with Rahul.

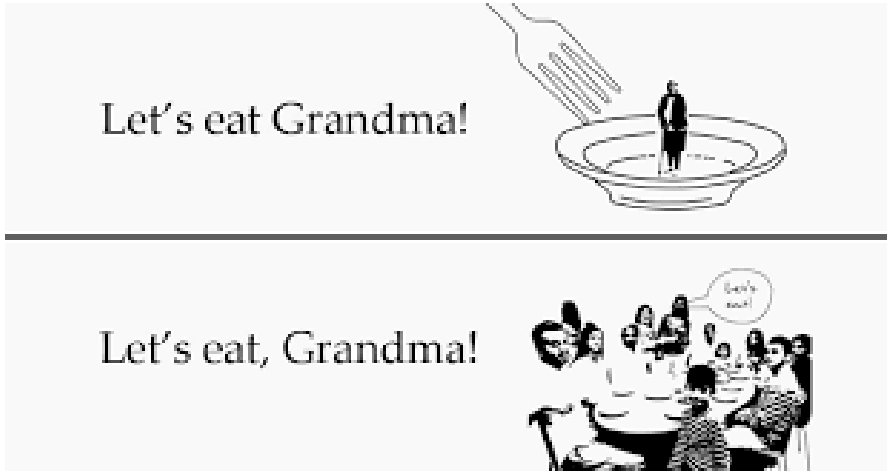
Similar surface structures but different interpretations!

Ambiguity and Punctuations!

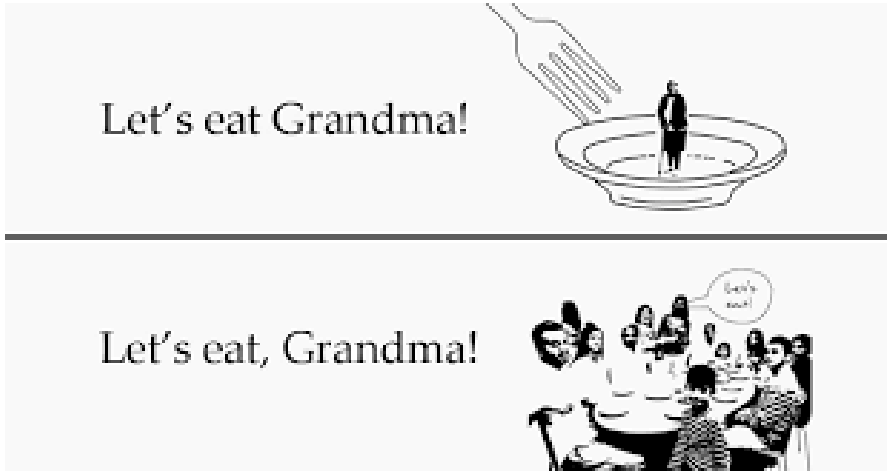


Ambiguity and Punctuations!

A woman without her man is nothing.



Ambiguity and Punctuations!



A woman without her man is nothing.

A woman, without her man, is nothing.

A woman: without her, man is nothing.

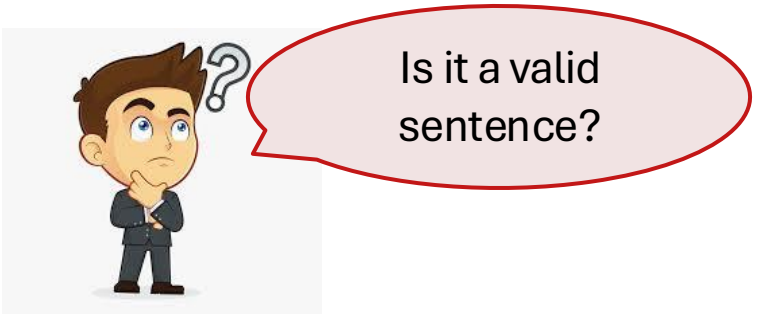
Punctuation is powerful.

What About This?

Buffalo buffalo Buffalo buffalo buffalo buffalo Buffalo buffalo

Dmitri Borgmann's *Beyond Language: Adventures in Word and Thought*. 1967.

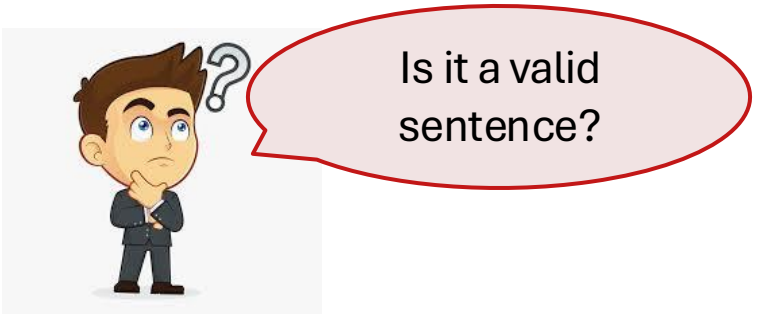
What About This?



Buffalo buffalo Buffalo buffalo buffalo buffalo Buffalo buffalo

Dmitri Borgmann's *Beyond Language: Adventures in Word and Thought*. 1967.

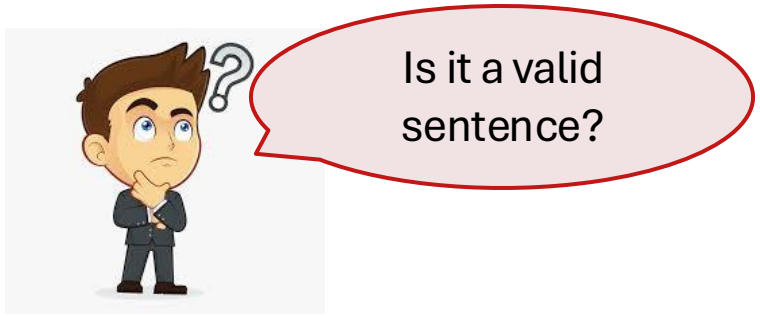
What About This?



Buffalo buffalo Buffalo buffalo buffalo buffalo Buffalo buffalo

Dmitri Borgmann's *Beyond Language: Adventures in Word and Thought*. 1967.

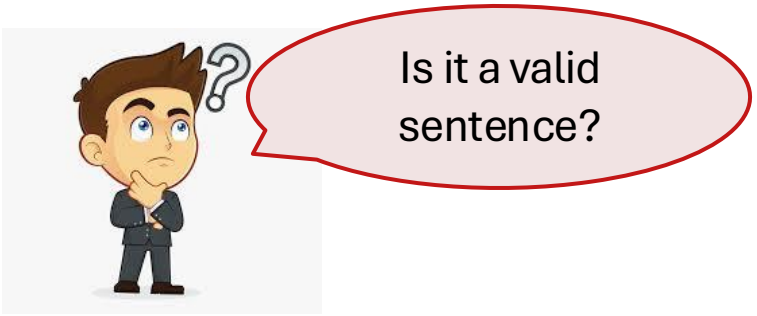
What About This?



Buffalo buffalo Buffalo buffalo buffalo buffalo Buffalo buffalo

Dmitri Borgmann's *Beyond Language: Adventures in Word and Thought*. 1967.

What About This?



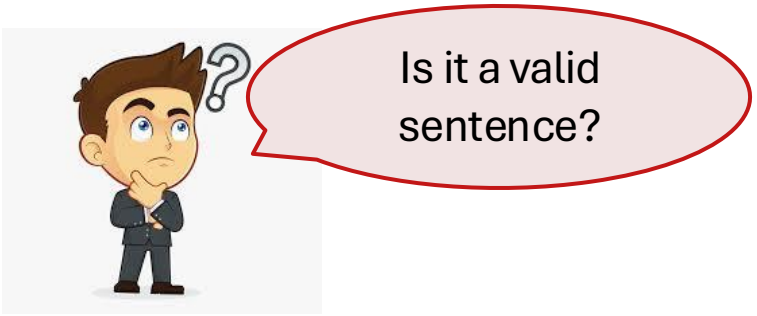
Buffalo buffalo Buffalo buffalo buffalo buffalo Buffalo buffalo

The word *buffalo* has three senses:

1. Noun: Animal (plural is also buffalo)
2. Proper Noun: American State
3. Verb: To bully someone

Dmitri Borgmann's *Beyond Language: Adventures in Word and Thought*. 1967.

What About This?



Buffalo buffalo Buffalo buffalo buffalo buffalo Buffalo buffalo

The word *buffalo* has three senses:

1. Noun: Animal (plural is also buffalo)
2. Proper Noun: American State
3. Verb: To bully someone

Buffalo buffalo, whom other Buffalo *buffalo* buffalo, buffalo Buffalo buffalo

Dmitri Borgmann's *Beyond Language: Adventures in Word and Thought*. 1967.

Why Else is Natural Language Understanding Difficult?

Non-standard English

Great job @justinbieber! Were SOO PROUD of what youve accomplished! U taught us 2 #neversaynever & you yourself should never give up either ♥

Segmentation Issues

the New York-New Haven Railroad

the New York-New Haven Railroad

Idioms / Multiword

dark horse
get cold feet
lose face
throw in the towel
Khana-wana (Echo)

Neologisms

unfriend
Retweet
bromance

World Knowledge

Mary and Juhi are sisters.
Mary and Juhi are mothers.

Tricky Entity Names

Where is *A Bug's Life* playing …
Let It Be was recorded …
… a mutation on the *for* gene …

Components of NLP

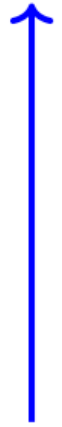
Natural Language Understanding

Natural Language Generation

NLP Layers

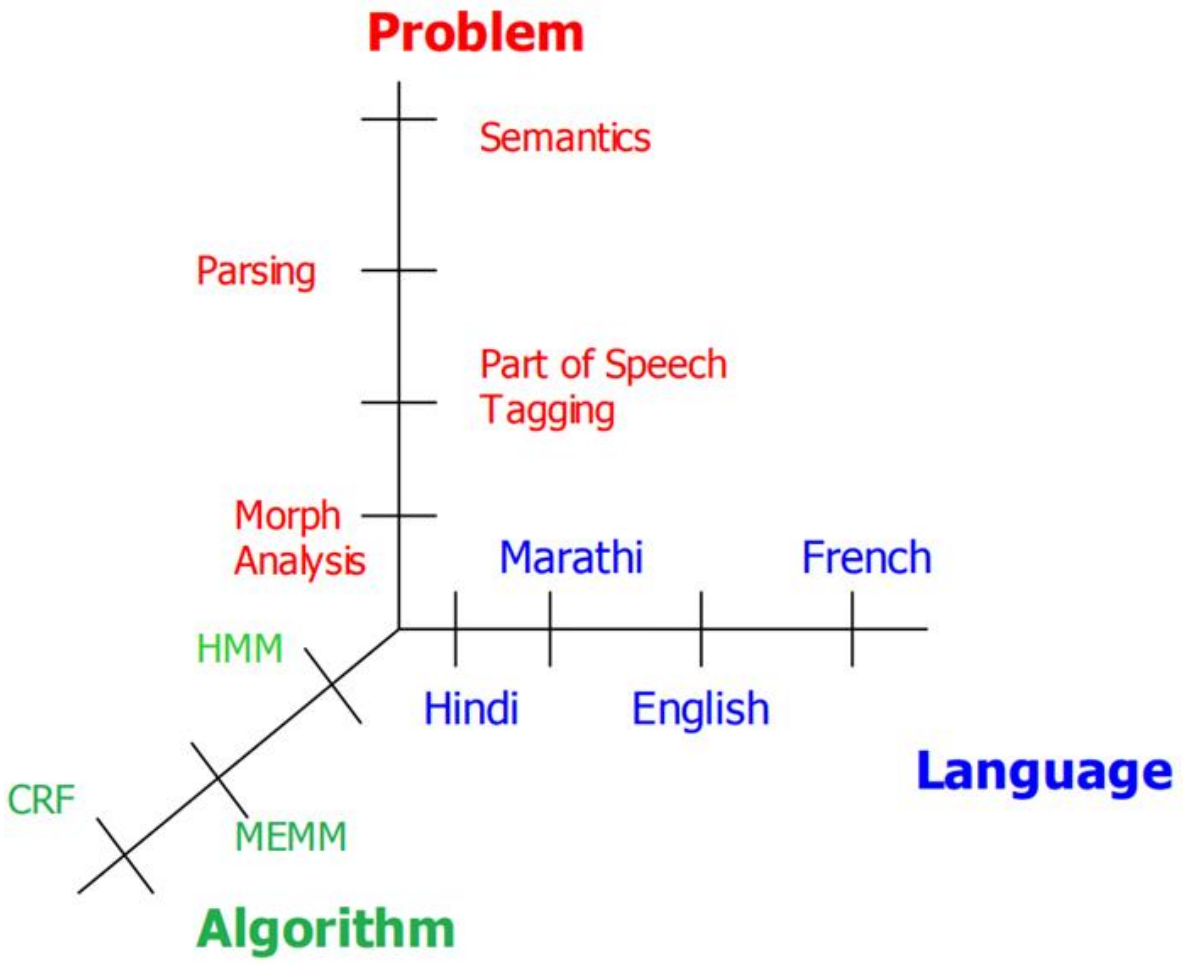
- Understanding the semantics is a non-trivial task.
- Needs to performs a series of incremental tasks to achieve this.
- NLP happens in layers.

Pragmatics & Discourse	*Study of semantics in context.*
Semantics	*Meaning of the sentence.*
Parsing	*Syntactic structure of the sentence.*
Chunking	*Grouping of meaningful phrases.*
Part of speech tagging	*Grammatical classes.*
Morphology	*Study of word structure.*



Increasing Complexity Of Processing

NLP Trinity


Problem
Semantics
Parsing
Part of Speech Tagging
Morph Analysis
Marathi
French
HMM
Hindi
English
Language
CRF
MEMM
Algorithm

Word and Token

Word: Smallest sequence of phonemes of a spoken language that can be uttered in isolation.

Word Segmentation/Tokenization: Breaking a string of characters into a sequence of words.

Token: Smallest sequence of graphemes that are delimited with some predefined characters (space, comma, full-stop, etc.);

Ram, Shyam, and Mohan are playing.	⇒	[Ram] [,] [Shyam] [,] [and] [Mohan] [are] [playing] [.]
21,53,010 COVID cases in India.	⇒	[21] [,] [53] [,] [010] [COVID] [cases] [in] [India] [.]
		[21,53,010] [COVID] [cases] [in] [India] [.] ✅
Check this out...https://www.abc.com	⇒	[Check] [this] [out] [.] [.] [.] [https] [:] [/] [/] [www] [.] [abc] [.] [com]
		[Check] [this] [out] [...] [https://www.abc.com] ✅
#GreatDayEver	⇒	[#] [Great] [Day] [Ever]

Morphology

Morphology is the study of words, how they are formed, and their relationship to other words in the same language. It analyzes the structure of words and parts of words, such as stems, root words, prefixes, and suffixes.

- Field of linguistics that studies the internal structure of words
  - How they are formed
  - Their relationship to other words in the same language.
- It defines word formation rule from the root word.
- *Morpheme* is the smallest linguistic unit that has semantic meaning
  - *Example:*
    - *"Pre", "ed", "ing", "s", "es", etc.*
      - Dogs ⇒ dog + s (plural)
      - Going ⇒ go + ing (present participle)
      - Independently ⇒ independent + ly (Adverb)
        ⇒ in + dependent + ly (Negation)
        ⇒ in + depend + ent + ly (relying)
        ⇒ in + de + pend + ent + ly

pend: (verb) to remain undecided or unsettled.

Morphology

- English, Chinese, etc. are commonly referred as *morphologically-poor* language.
- Hindi, Turkish, Hungarian, etc. are termed as *morphologically-rich* language.

English	Hindi	Linguistic property
I will go.	मैं जाऊँगा।	Different morphological forms of word 'will go' in Hindi
We will go.	हम जाएंगे।	
You will go.	तुम जाओगे।	
He will go.	वह जाएगा।	
She will go.	वह जाएगी।	

Syntax

Syntax concerns the way in which words can be combined together to form (grammatical) sentences.

Syntax

Syntax concerns the way in which words can be combined together to form (grammatical) sentences.

Pragmatics & Discourse	*Study of semantics in context.*
Semantics	*Meaning of the sentence.*
Parsing	*Syntactic structure of the sentence.*
Chunking	*Grouping of meaningful phrases.*
Part of speech tagging	*Grammatical classes.*
Morphology	*Study of word structure.*

Increasing Complexity Of Processing

Parts-of-Speech (POS)

Grammatical class of the word.

Parts-of-Speech (POS)

Grammatical class of the word.

He	ate	an	apple	.

Parts-of-Speech (POS)

Grammatical class of the word.

He	ate	an	apple	.
PRP	VBD	DT	NN	.

Tags

PRP: Personal Pronoun
VBD: Verb, Past
DT: Determiner
NN: Noun, Singular, Mass
TO: *to*
IN: Preposition

- 45 tags in Penn Treebank tagset
- 146 tags in C7

Parts-of-Speech (POS)

Grammatical class of the word.

He	ate	an	apple	.
PRP	VBD	DT	NN	.

PoS disambiguation:

- A word can belong to different grammatical classes.

Tags

PRP: Personal Pronoun
VBD: Verb, Past
DT: Determiner
NN: Noun, Singular, Mass
TO: *to*
IN: Preposition

- 45 tags in Penn Treebank tagset
- 146 tags in C7

Parts-of-Speech (POS)

Grammatical class of the word.

He	ate	an	apple	.
PRP	VBD	DT	NN	.

Tags

PRP: Personal Pronoun
VBD: Verb, Past
DT: Determiner
NN: Noun, Singular, Mass
TO: *to*
IN: Preposition

- 45 tags in Penn Treebank tagset
- 146 tags in C7

PoS disambiguation:

- A word can belong to different grammatical classes.

He	went	to	the	***park***	in	a	car	.
PRP	VBD	TO	DT	***NN***	IN	DT	NN	.

They	went	to	***park***	the	car	in	the	shed	.
PRP	VBD	TO	***VB***	DT	NN	IN	DT	NN	.

Chunking

Identification of non-recursive phrases (noun, verb, etc.)

- He went to the Indian city Mumbai. ⇒
  [NP He] [VP went] [PP to] [NP the Indian city Mumbai]

- Mumbai green lights women icons on traffic signals earns global praise. ⇒
  [NP Mumbai green lights women icons] [PP on] [NP traffic signals] [VP earns] [NP global praise]

Syntax Processing

Validate the grammatical structure of the sentence.

Syntax Processing

Validate the grammatical structure of the sentence.

Let, vocabulary = [the, mango, he, eats, …]

He eats a mango. ⇒ ✅

He mango eats a. ⇒ ❌

- The sequence of words must follow the grammatical structure of the language to form a valid sentence.
  - Construct a parse tree.

Syntax Processing

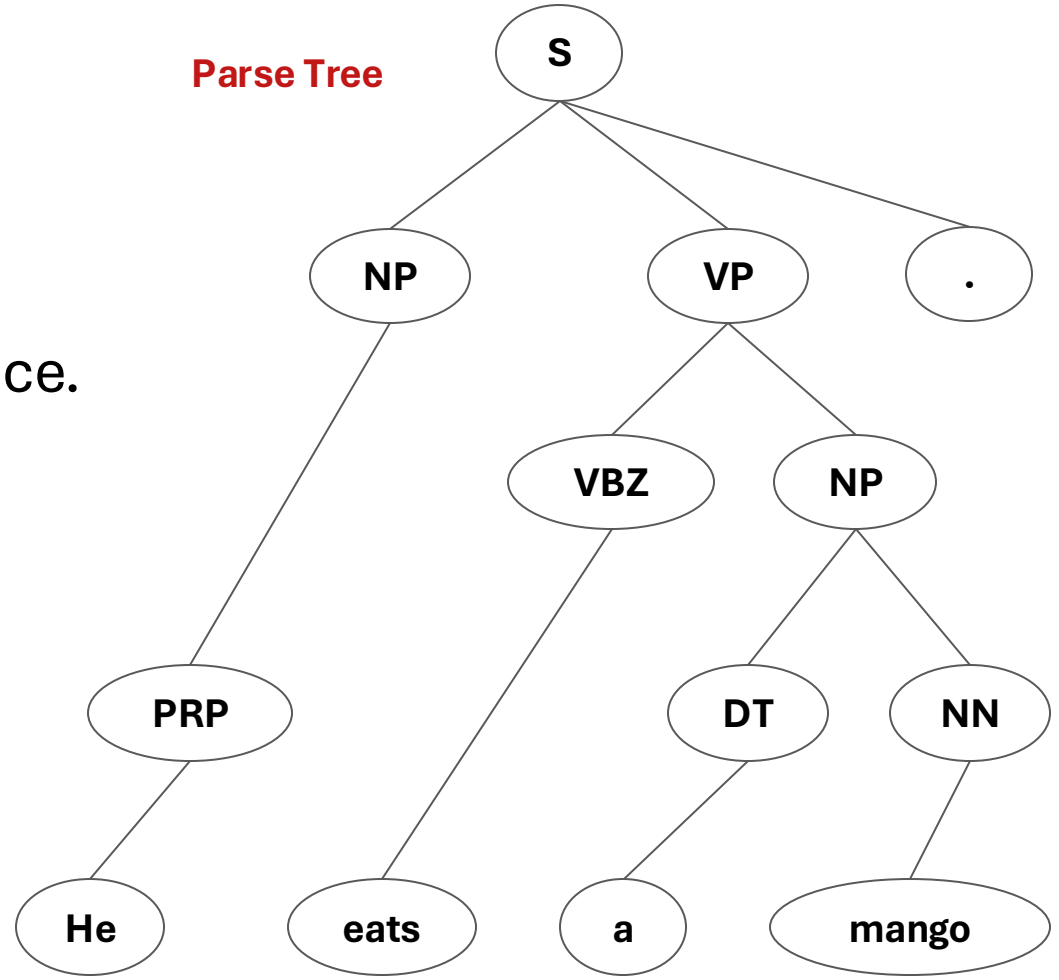
Validate the grammatical structure of the sentence.

Let, vocabulary = [the, mango, he, eats, …]

He eats a mango. ⇒ ✅

He mango eats a. ⇒ ❌

- The sequence of words must follow the grammatical structure of the language to form a valid sentence.
  - Construct a parse tree.



Syntax Processing

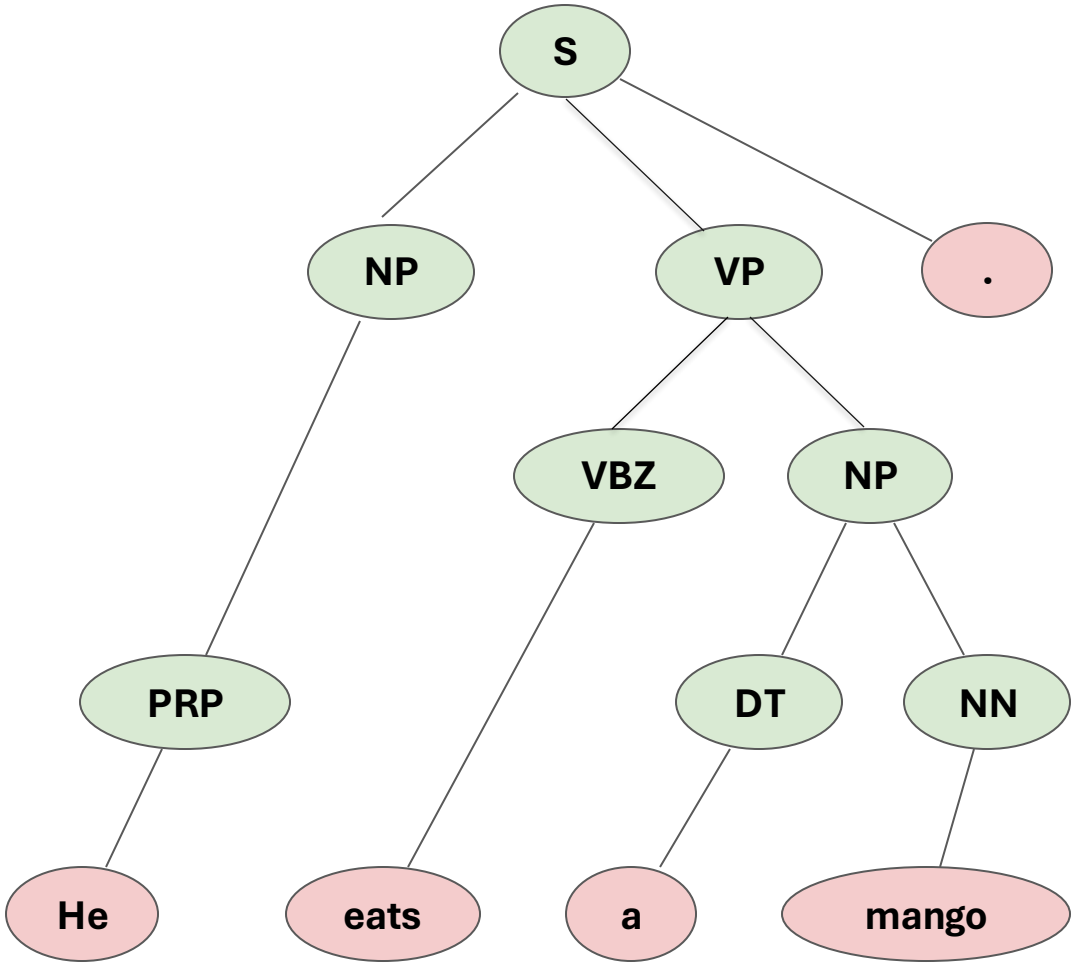
Every language has a grammar G = <V, T, P, S>.

Syntax Processing

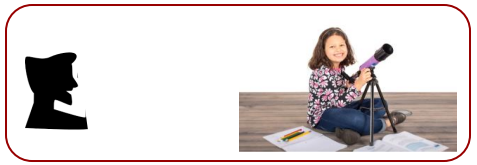
Every language has a grammar G = <V, T, P, S>.

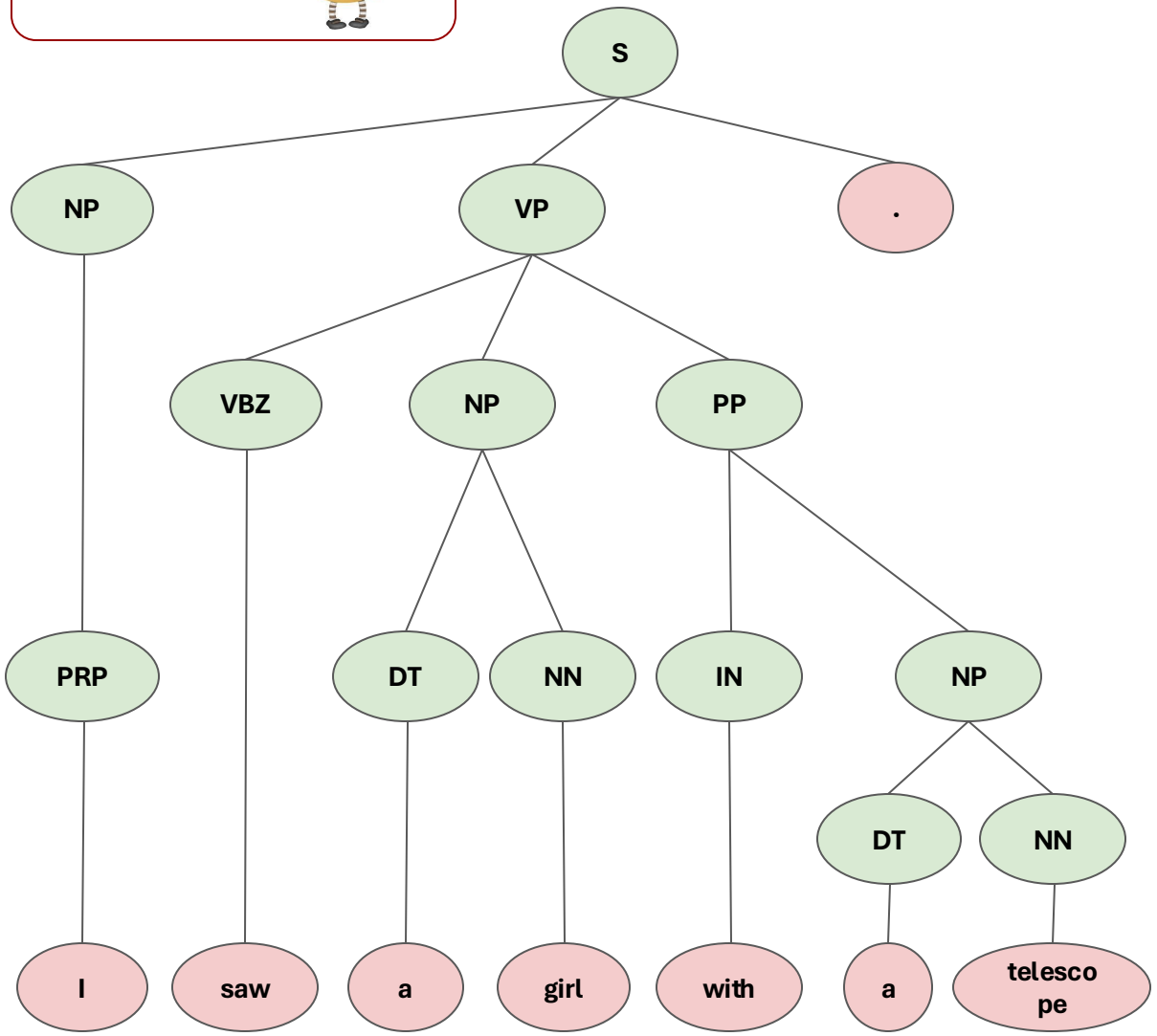
Productions (P) or rules:

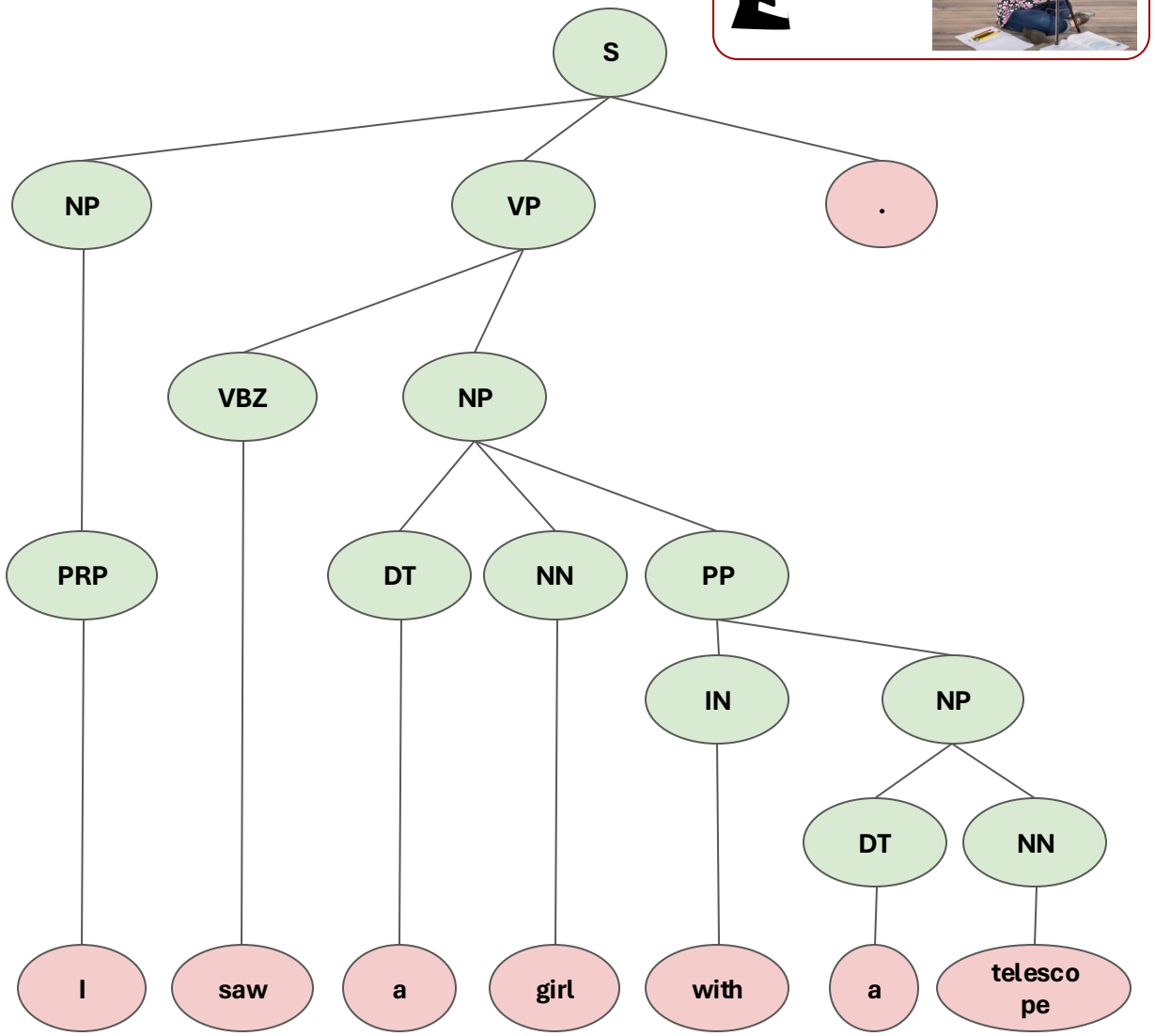
S	→	NP VP .
NP	→	PRP | NN | DT NP
VP	→	VBZ NP
PRP	→	He
VBZ	→	eats
DT	→	a
NN	→	mango



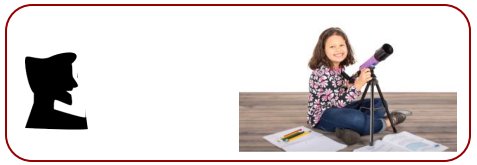
Syntactic Ambiguity

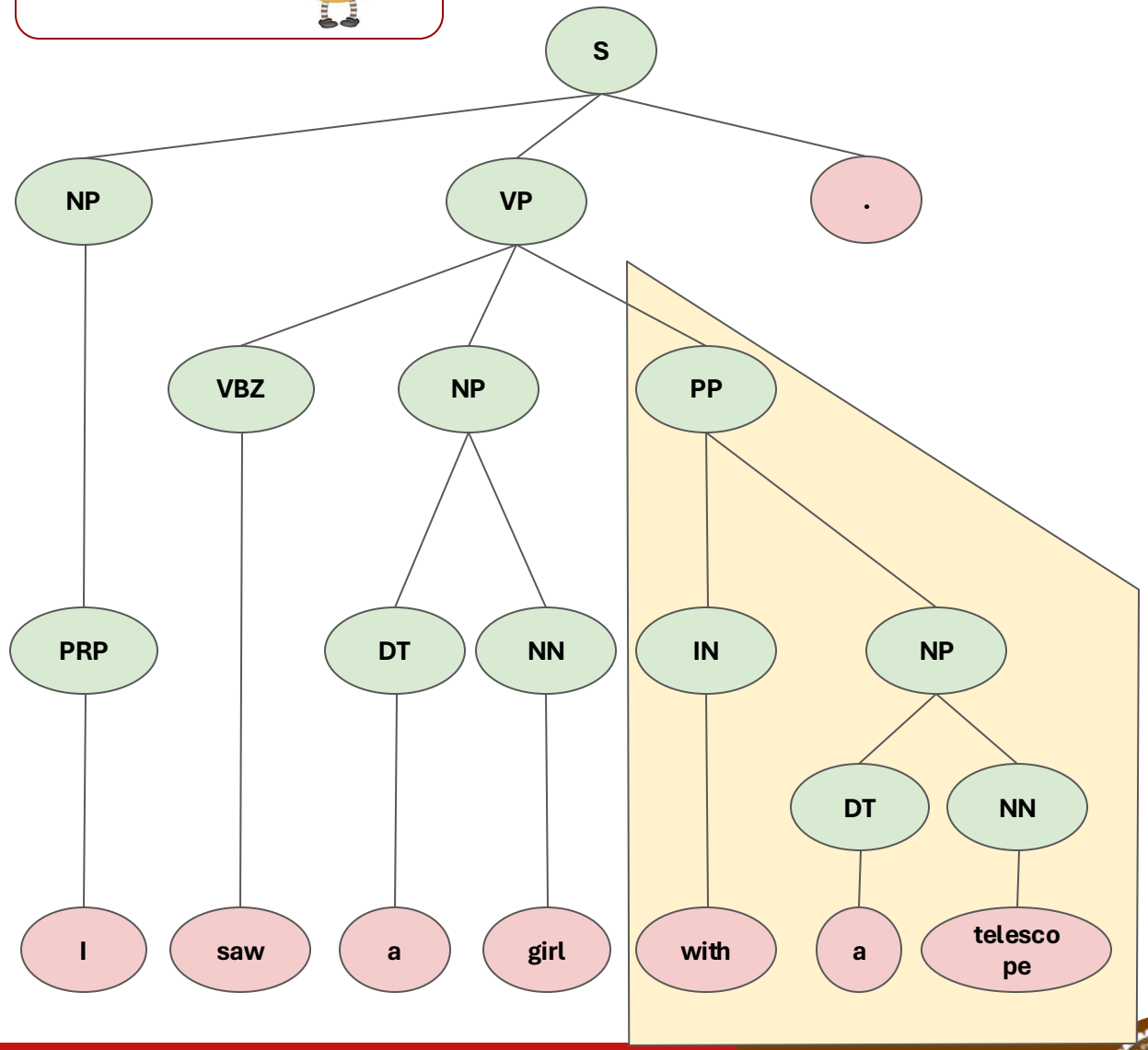


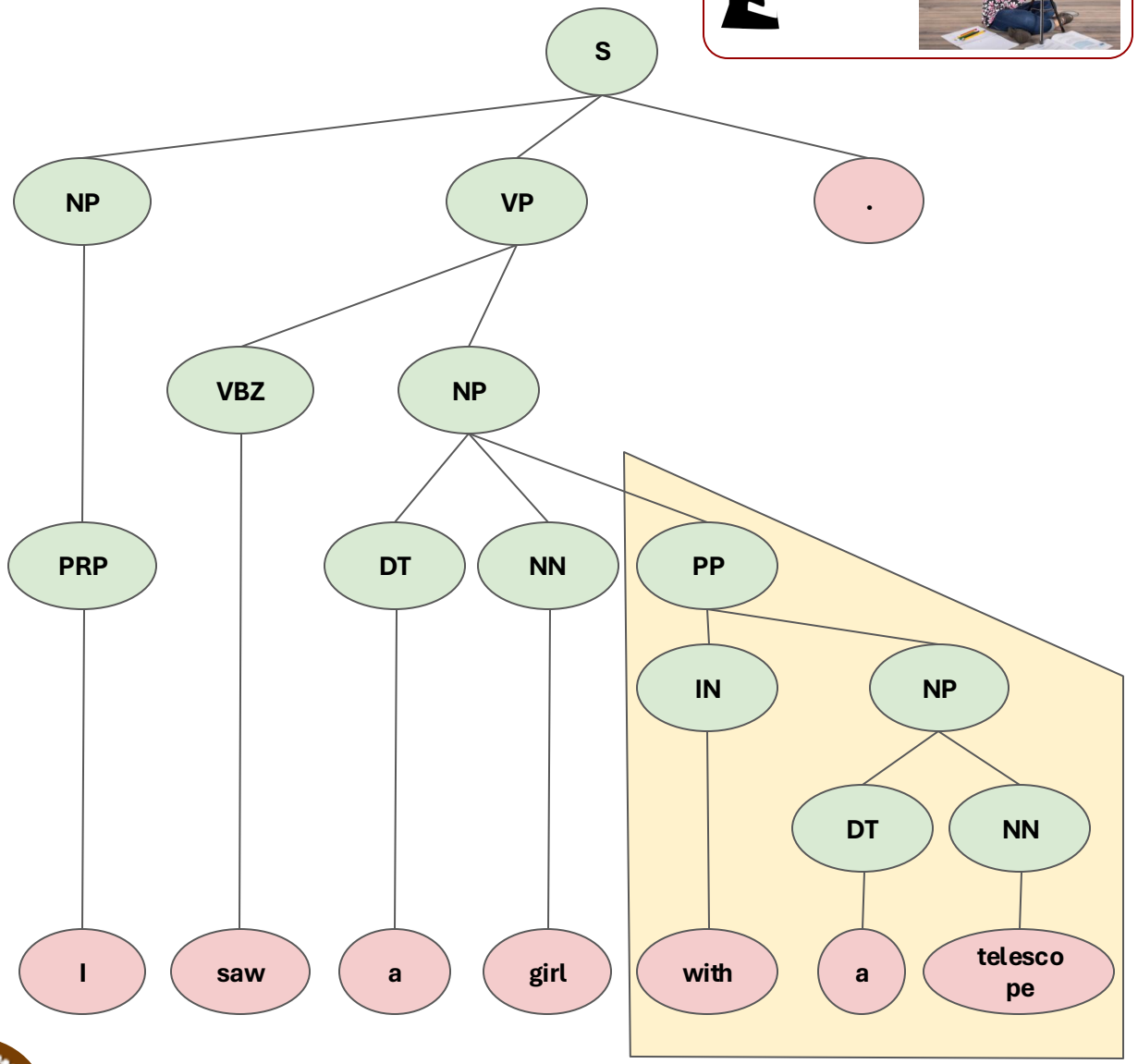




Syntactic Ambiguity







Semantics

Semantics (and pragmatics) are the glue that connect language to the real world.

Semantics

Semantics (and pragmatics) are the glue that connect language to the real world.

Semantics is concerned with the meaning of words and how to combine words into meaningful phrases and sentences.

Semantics

Semantics (and pragmatics) are the glue that connect language to the real world.

Semantics is concerned with the meaning of words and how to combine words into meaningful phrases and sentences.

- **Decompositional** – What the "components" of meaning "in" a word are
- **Ontological** – How the meaning of the word relates to the meanings of other words
- **Distributional** – What contexts the word is found in, relative to other words

Decompositional Semantics

Decompositional Semantics Divides the Meaning of Words into Components

What are its strengths and weaknesses?

boy [+human −female −adult]	girl [+human +female −adult]
man [+human −female +adult]	woman [+human +female +adult]

Ontological Semantics

Ontological semantics says that the meaning of a word is its relationship to other words.

Ontological Semantics

Ontological semantics says that the meaning of a word is its relationship to other words.

The Basic (Ontological) Semantic Relations

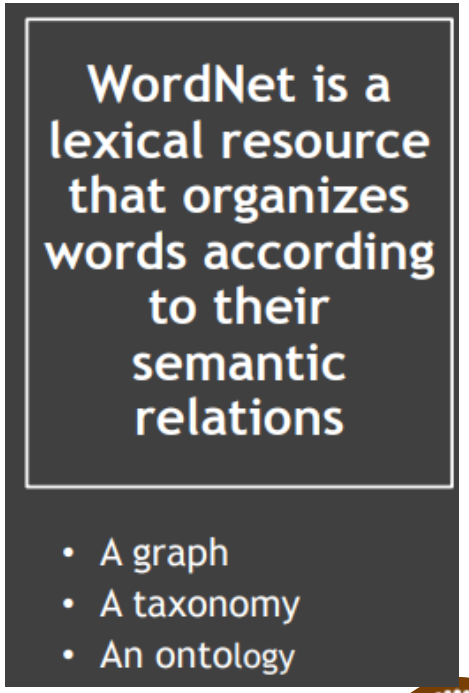
- Synonymy—equivalence
  - <small, little>
- Antonymy—opposition
  - <small, large>
- Hyponymy—subset; is-a relation
  - <dog, mammal>
- Hypernymy—superset
  - <mammal, dog>
- Meronymy—part-of relation
  - <liver, body>
- Holonymy—has-a relation
  - <body, liver>

Ontological Semantics

Ontological semantics says that the meaning of a word is its relationship to other words.

The Basic (Ontological) Semantic Relations

- Synonymy—equivalence
  - <small, little>
- Antonymy—opposition
  - <small, large>
- Hyponymy—subset; is-a relation
  - <dog, mammal>
- Hypernymy—superset
  - <mammal, dog>
- Meronymy—part-of relation
  - <liver, body>
- Holonymy—has-a relation
  - <body, liver>



Ontological Semantics

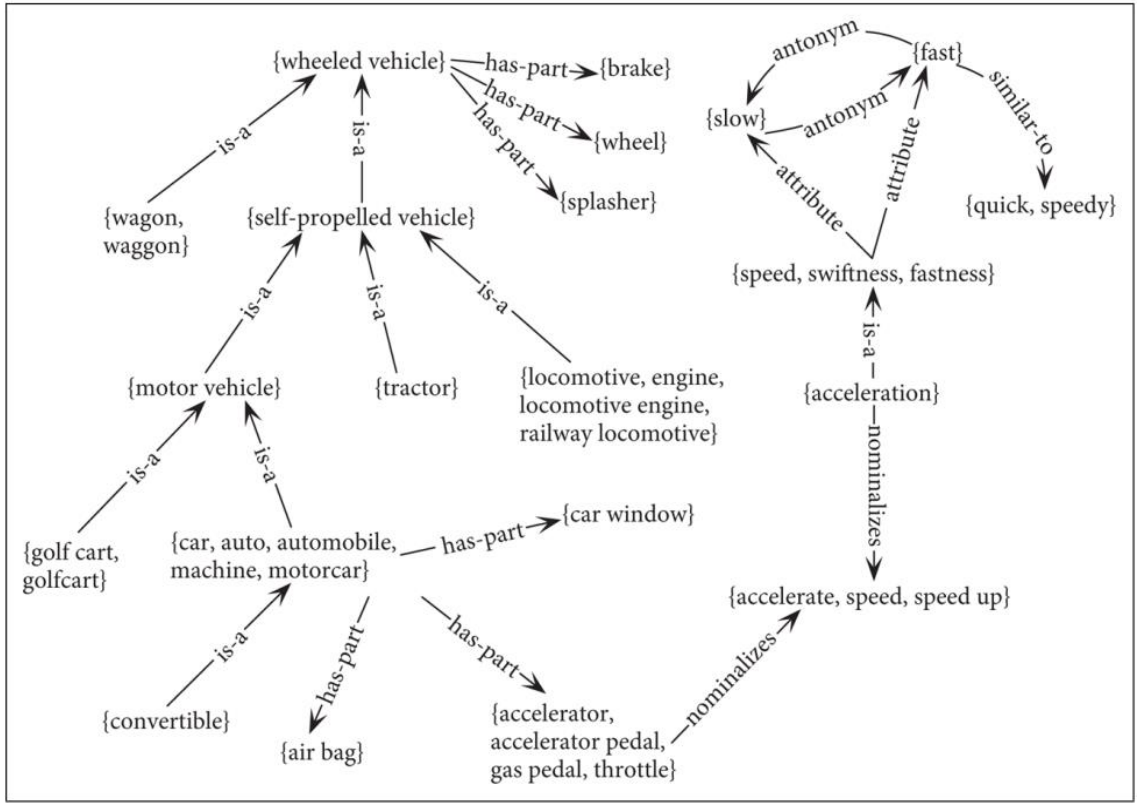
Ontological semantics says that the meaning of a word is its relationship to other words.

The Basic (Ontological) Semantic Relations

- Synonymy—equivalence
  - <small, little>
- Antonymy—opposition
  - <small, large>
- Hyponymy—subset; is-a relation
  - <dog, mammal>
- Hypernymy—superset
  - <mammal, dog>
- Meronymy—part-of relation
  - <liver, body>
- Holonymy—has-a relation
  - <body, liver>

WordNet is a lexical resource that organizes words according to their semantic relations

- A graph
- A taxonomy
- An ontology



Distributional Semantics

The meanings of words can be derived from their distributional properties in large corpora of text. It relies on the context in which words appear.

Distributional Semantics

The meanings of words can be derived from their distributional properties in large corpora of text. It relies on the context in which words appear.

Example: The meaning of the word "cat" can be inferred from the contexts it appears in, such as sentences where it co-occurs with words like "pet," "animal," "meow," and "feline."

The co-occurrence matrix

	leash	walk	run	owner	pet	bark
dog	3	5	2	5	3	2
cat	0	3	3	2	3	0
lion	0	3	2	0	1	0
light	0	0	0	0	0	0
bark	1	0	0	2	1	0
car	0	0	1	3	0	0

Pragmatics

Pragmatics considers [Thomas, 1995]:

- the negotiation of meaning between speaker and listener.
- the context of the utterance.
- the intention of the user.

Pragmatics

Pragmatics considers [Thomas, 1995]:

- the negotiation of meaning between speaker and listener.
- the context of the utterance.
- the intention of the user.

- Context/World knowledge: An employee coming late to the office.
  - Utterance: Do you know what time is it?
  - Literal meaning: Are you aware of the current time? (Response: Yes, it is 12:30 PM)
  - Pragmatic meaning: Why are you coming so late? (Response: Reason for being late.)

Pragmatics

Pragmatics considers [Thomas, 1995]:

- the negotiation of meaning between speaker and listener.
- the context of the utterance.
- the intention of the user.

- Context/World knowledge: An employee coming late to the office.
  - Utterance: Do you know what time is it?
  - Literal meaning: Are you aware of the current time? (Response: Yes, it is 12:30 PM)
  - Pragmatic meaning: Why are you coming so late? (Response: Reason for being late.)
- Intention:
  - Utterance: Can you pass the water bottle?
  - Literal meaning: Are you able to pass the water bottle? (Response: Yes, I can.)
  - Pragmatic meaning: Pass me the water bottle. (Response: Handover the water bottle)

Discourse

Processing of sequence of sentences.

Discourse

Processing of sequence of sentences.

Mother said to John: Go to school. It is open today. Are you planning to bunk? Father will be very angry.

Discourse

Processing of sequence of sentences.

Mother said to John: Go to school. It is open today. Are you planning to bunk? Father will be very angry.

Discourse processing helps answering these questions:

- What is open?
- Bunk what?
- Why the father will be angry?

Tasks we want to solve in NLP

Semantic Role Labelling (SRL)

Semantic Role Labelling (SRL)

Identify the semantic role of each argument (noun phrase) w.r.t. the predicate (main verb) of the sentence.

Semantic Role Labelling (SRL)

Identify the semantic role of each argument (noun phrase) w.r.t. the predicate (main verb) of the sentence.

John	**drove**	Mary	from	Delhi	to	Pune	in	his	car
Agent		Patient		source		destination			instrument

Ram	**hit**	Shyam	with	a	hockey	stick	yesterday
Agent		Patient			instrument		time

Textual Entailment

Textual Entailment

Determine whether one natural language sentence entails (implies) another under an ordinary interpretation.

Textual Entailment

Determine whether one natural language sentence entails (implies) another under an ordinary interpretation.

(*Ram hit Shyam with a hockey stick yesterday*. → *Shyam got hurt*) ⇒ Positive TE

(*Ram hit Shyam with a hockey stick yesterday*. → *Shyam did not get hurt*) ⇒ Negative TE

(*Ram hit Shyam with a hockey stick yesterday*. → *Shyam got his first goal*) ⇒ Non TE

Co-reference Resolution

Co-reference Resolution

- Two referring expressions used to refer to the same entity are said to **co-refer**.
- Determine which phrases in a document co-refer.

Co-reference Resolution

- Two referring expressions used to refer to the same entity are said to **co-refer**.
- Determine which phrases in a document co-refer.

John shows Bob his Toyota yesterday. It's similar to the one I bought five years ago.

That was really nice, but he likes this *one* even better.

Co-reference Resolution

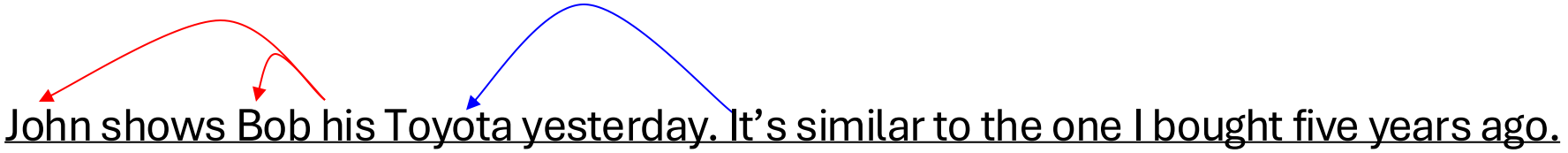
- Two referring expressions used to refer to the same entity are said to **co-refer**.
- Determine which phrases in a document co-refer.

John shows Bob his Toyota yesterday. It's similar to the one I bought five years ago.

That was really nice, but he likes this *one* even better.

Co-reference Resolution

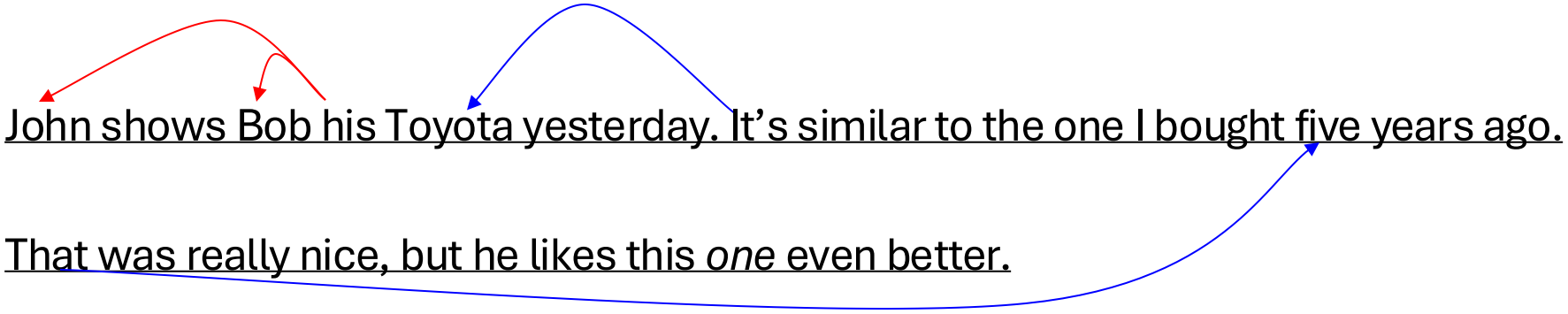
- Two referring expressions used to refer to the same entity are said to **co-refer**.
- Determine which phrases in a document co-refer.



That was really nice, but he likes this *one* even better.

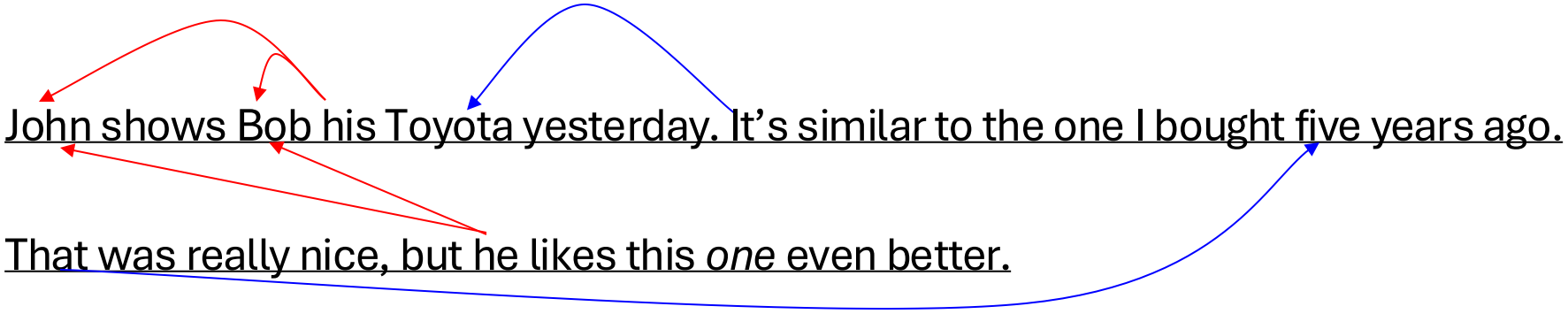
Co-reference Resolution

- Two referring expressions used to refer to the same entity are said to **co-refer**.
- Determine which phrases in a document co-refer.



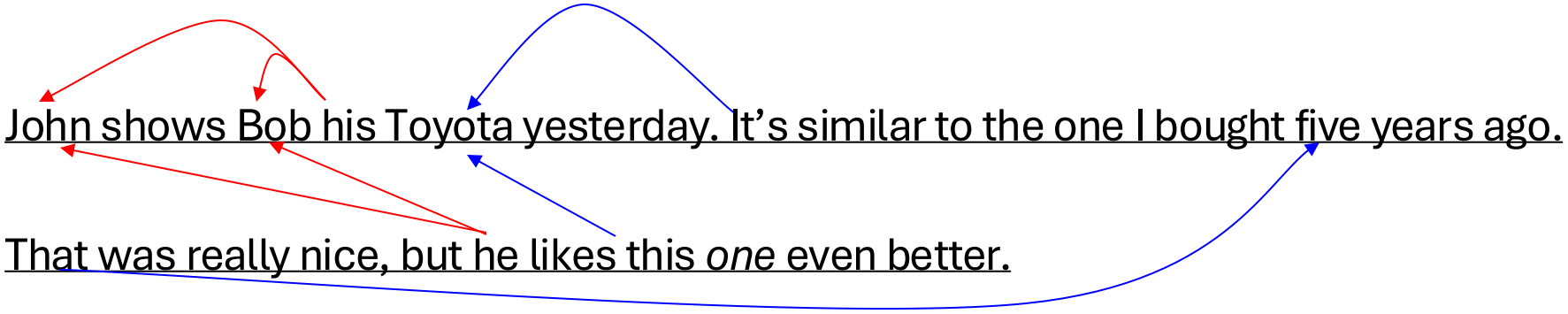
Co-reference Resolution

- Two referring expressions used to refer to the same entity are said to **co-refer**.
- Determine which phrases in a document co-refer.



Co-reference Resolution

- Two referring expressions used to refer to the same entity are said to **co-refer**.
- Determine which phrases in a document co-refer.



Information Extraction

Information Extraction

Extraction of relevant piece of information.

Information Extraction

Extraction of relevant piece of information.

- Named Entity Recognition (NER):
  - Identify names (Proper nouns)
    - [India]$_{\text{Location}}$ born [Sundar Pichai]$_{\text{Person}}$ is the CEO of [Google]$_{\text{Organization}}$ and its parent company [Alphabet]$_{\text{Organization}}$

Information Extraction

Extraction of relevant piece of information.

- Named Entity Recognition (NER):
  - Identify names (Proper nouns)
    - [India]$_{\text{Location}}$ born [Sundar Pichai]$_{\text{Person}}$ is the CEO of [Google]$_{\text{Organization}}$ and its parent company [Alphabet]$_{\text{Organization}}$

- Relation Extraction:
  - Relation among entities
    - CEO(Sundar Pichai, Google), CEO(Sundar Pichai, Alphabet), Born-at(Sundar Pichai, India), ParentOrg(Alphabet, Google)

Word Sense Disambiguation (WSD)

Word Sense Disambiguation (WSD)

What does a word mean?

Word Sense Disambiguation (WSD)

What does a word mean?

- The fisherman went to the ***bank***. ⇒ Financial bank or river bank?

Word Sense Disambiguation (WSD)

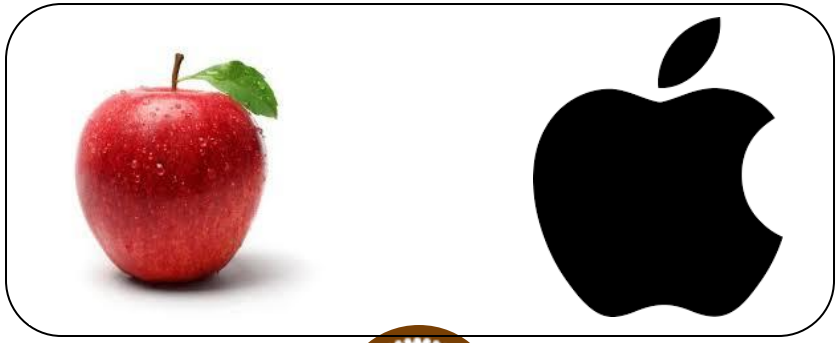
What does a word mean?

- The fisherman went to the ***bank***. ⇒ Financial bank or river bank?
    - The fisherman went to the ***bank*** to withdraw money.
    - The fisherman went to the ***bank*** to fish.

Word Sense Disambiguation (WSD)

What does a word mean?

- The fisherman went to the ***bank***. ⇒ Financial bank or river bank?
    - The fisherman went to the ***bank*** to withdraw money.
    - The fisherman went to the ***bank*** to fish.



Sentiment Analysis

Extract polarity orientation of the subjectivity.

- Really superb pillow. Love to sleep on it.. very comfortable… ⇒ Positive
- It's a mass Chinese product. Too expensive. Thin and useless ⇒ Negative
- My neighbours are home and it’s good to wake up at 3am in the morning. ⇒ Negative?
- Campus has deadly snakes. ⇒ Negative
- Shane Warne is a deadly spinner. ⇒ Positive?
- The food was cheap. ⇒ Positive?
- Not to mention the cheap service I got at the restaurant. ⇒ Negative
- Movie was 4 hours long. ⇒ Neutral?

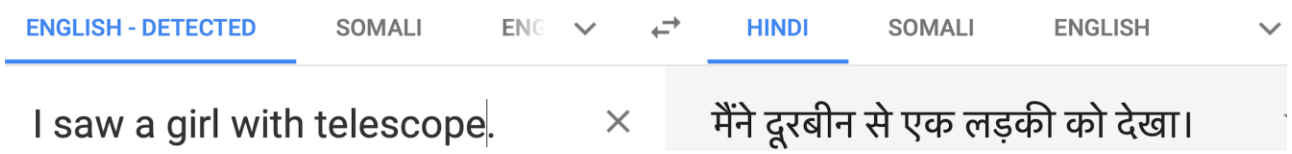
Machine Translation

Machine Translation

Given a sentence in the source language L1, convert it to the target language L2, such that the semantic (adequacy and fluency) is preserved.

Machine Translation

Given a sentence in the source language L1, convert it to the target language L2, such that the semantic (adequacy and fluency) is preserved.



Source: Google Translate

Machine Translation

Given a sentence in the source language L1, convert it to the target language L2, such that the semantic (adequacy and fluency) is preserved.

Source: Google Translate

Machine Translation

Given a sentence in the source language L1, convert it to the target language L2, such that the semantic (adequacy and fluency) is preserved.

Source: Google Translate

Machine Translation

Given a sentence in the source language L1, convert it to the target language L2, such that the semantic (adequacy and fluency) is preserved.

Source: Google Translate

Summarization

Summarization

Given a document, summarize the semantics (extract relevant information) in shorter length text.

Summarization

Given a document, summarize the semantics (extract relevant information) in shorter length text.

Document

Sen. Barack Obama sealed the Democratic presidential nomination last night after a grueling and history-making campaign against Sen. Hillary Rodham Clinton that will make him the first African American to head a major-party ticket.

Summarization

Given a document, summarize the semantics (extract relevant information) in shorter length text.

Document

Sen. Barack Obama sealed the Democratic presidential nomination last night after a grueling and history-making campaign against Sen. Hillary Rodham Clinton that will make him the first African American to head a major-party ticket.

Summary

Barack Obama is the Democratic presidential candidate.

Question Answering

Question Answering

Answer natural language questions based on information presented in the repository.

Question Answering

Answer natural language questions based on information presented in the repository.

Factoid Questions

- Question: Who is the author of the book Wings of Fire?
- Answer: A. P. J. Abdul Kalam

Question Answering

Answer natural language questions based on information presented in the repository.

Factoid Questions

- Question: Who is the author of the book Wings of Fire?
- Answer: A. P. J. Abdul Kalam

List Questions

- Question: What are the islands in India?
- Answer: Andaman Island, Nicobar Island, Labyrinth Island, Barren Island

Question Answering

Answer natural language questions based on information presented in the repository.

Factoid Questions

- Question: Who is the author of the book Wings of Fire?
- Answer: A. P. J. Abdul Kalam

List Questions

- Question: What are the islands in India?
- Answer: Andaman Island, Nicobar Island, Labyrinth Island, Barren Island

Descriptive Questions

- Question: What is Greenhouse effect?
- Answer: The analogy used to describe the ability of gases in the atmosphere to absorb heat from the earth's surface.

Dialog System and Chatbot

Dialog System and Chatbot

C_1:	. . . I need to travel in May.
A_1:	And, what day in May did you want to travel?
C_2:	OK uh I need to be there for a meeting that's from the 12th to the 15th.
A_2:	And you're flying into what city?
C_3:	Seattle.
A_3:	And what time would you like to leave Pittsburgh?
C_4:	Uh hmm I don't think there's many options for non-stop.
A_4:	Right. There's three non-stops today.
C_5:	What are they?
A_5:	The first one departs PGH at 10:00am arrives Seattle at 12:05 their time. The second flight departs PGH at 5:55pm, arrives Seattle at 8pm. And the last flight departs PGH at 8:15pm arrives Seattle at 10:28pm.
C_6:	OK I'll take the 5ish flight on the night before on the 11th.
A_6:	On the 11th? OK. Departing at 5:55pm arrives Seattle at 8pm, U.S. Air flight 115.
C_7:	OK.

Dialog System and Chatbot

C_1:	. . . I need to travel in May.
A_1:	And, what day in May did you want to travel?
C_2:	OK uh I need to be there for a meeting that's from the 12th to the 15th.
A_2:	And you're flying into what city?
C_3:	Seattle.
A_3:	And what time would you like to leave Pittsburgh?
C_4:	Uh hmm I don't think there's many options for non-stop.
A_4:	Right. There's three non-stops today.
C_5:	What are they?
A_5:	The first one departs PGH at 10:00am arrives Seattle at 12:05 their time. The second flight departs PGH at 5:55pm, arrives Seattle at 8pm. And the last flight departs PGH at 8:15pm arrives Seattle at 10:28pm.
C_6:	OK I'll take the 5ish flight on the night before on the 11th.
A_6:	On the 11th? OK. Departing at 5:55pm arrives Seattle at 8pm, U.S. Air flight 115.
C_7:	OK.

User:	Do you like Doctor Who?
Cleverbot:	Yes, so funny.
User:	What's your favorite show?
Cleverbot:	I don't have one.
User:	What's your name?
Cleverbot:	Jason I already told you.
User:	No your real name.
Cleverbot:	Stephanie IS my real name.
User:	I thought it was Jason?
Cleverbot:	How did you come to that conclusion.
User:	Because you said so.
Cleverbot:	I have to do my homework.

Dialog System and Chatbot

C_1:	. . . I need to travel in May.
A_1:	And, what day in May did you want to travel?
C_2:	OK uh I need to be there for a meeting that's from the 12th to the 15th.
A_2:	And you're flying into what city?
C_3:	Seattle.
A_3:	And what time would you like to leave Pittsburgh?
C_4:	Uh hmm I don't think there's many options for non-stop.
A_4:	Right. There's three non-stops today.
C_5:	What are they?
A_5:	The first one departs PGH at 10:00am arrives Seattle at 12:05 their time. The second flight departs PGH at 5:55pm, arrives Seattle at 8pm. And the last flight departs PGH at 8:15pm arrives Seattle at 10:28pm.
C_6:	OK I'll take the 5ish flight on the night before on the 11th.
A_6:	On the 11th? OK. Departing at 5:55pm arrives Seattle at 8pm, U.S. Air flight 115.
C_7:	OK.

User:	Do you like Doctor Who?
Cleverbot:	Yes, so funny.
User:	What's your favorite show?
Cleverbot:	I don't have one.
User:	What's your name?
Cleverbot:	Jason I already told you.
User:	No your real name.
Cleverbot:	Stephanie IS my real name.
User:	I thought it was Jason?
Cleverbot:	How did you come to that conclusion.
User:	Because you said so.
Cleverbot:	I have to do my homework.

We will explore
how LLMs solve these tasks!