



Deep Learning for NLP (ELL 884)



Neuro-linguistic
programming

Deep Learning to NLP (ELL 884)



~~Neuro-linguistic
programming~~

Deep Learning to NLP (ELL 884)



**Non-Linear
Programming**

Deep Learning to NLP (ELL 884)



~~Non-Linear~~
~~Programming~~

Deep Learning to NLP (ELL 884)



Natural Language Processing

Deep Learning to NLP (ELL 884)

NLP (Wiki)

1. Natural Language Processing
2. Natural-linear Programming
3. Neuro-linguistic Programming
4. Natural-language Programming
5. National Library of Poland
6. National Library of the Philippines
7. No light perception
8. National Labour Party
9. National Liberal Party
10. National Liberation Party
11. Natural Law Party
12. New Labour Party

- **Course Instructor:** Tanmoy Chakraborty (NLP, Social Computing)
tanchak@iitd.ac.in
- **Guest Lecture:** TBD
- **Course page:** <https://lcs2.in/nlp2402>
- **Piazza:** http://piazza.com/iit_delhi/winter2025/deeplearningfornaturallanguageprocessing
[Code: **rd6ikjkzp7m**]
- **TAs:**
 - Sahil Mishra*, Aswini Kumar Padhi, Anwoy Chatterjee, Vaibhav Seth
- **Group Email:** TBD

Useful resources/tools/libraries

- Natural Language Toolkit (NLTK)
- Stanford CoreNLP
- CMU ARK for Noisy Text
- Scikit-learn
- Spacy
- Stanza
- Shallow Parser - for Indian Language
- Universal Parser - Multi-lingual
- HuggingFace

Prerequisite

- Excitement about language!
- Willingness to learn

Mandatory	Desirable
<ul style="list-style-type: none">• Data Structures & Algorithm• Machine Learning• Python programming	Deep learning

- Strongly recommended to learn ML. This class will not cover fundamentals of ML.

Course Directives

- **Class Time:** Mon & Thu, 2 pm – 3:30 pm
- **Office Hour:** Mon 5-6 pm
- **Room:** LH-308

HashLearn

- Meet your instructor at least once per 15 days to resolve your doubts.
- Mon 5-5:30 pm (**appointment based, email me at least 1 hr before coming**)

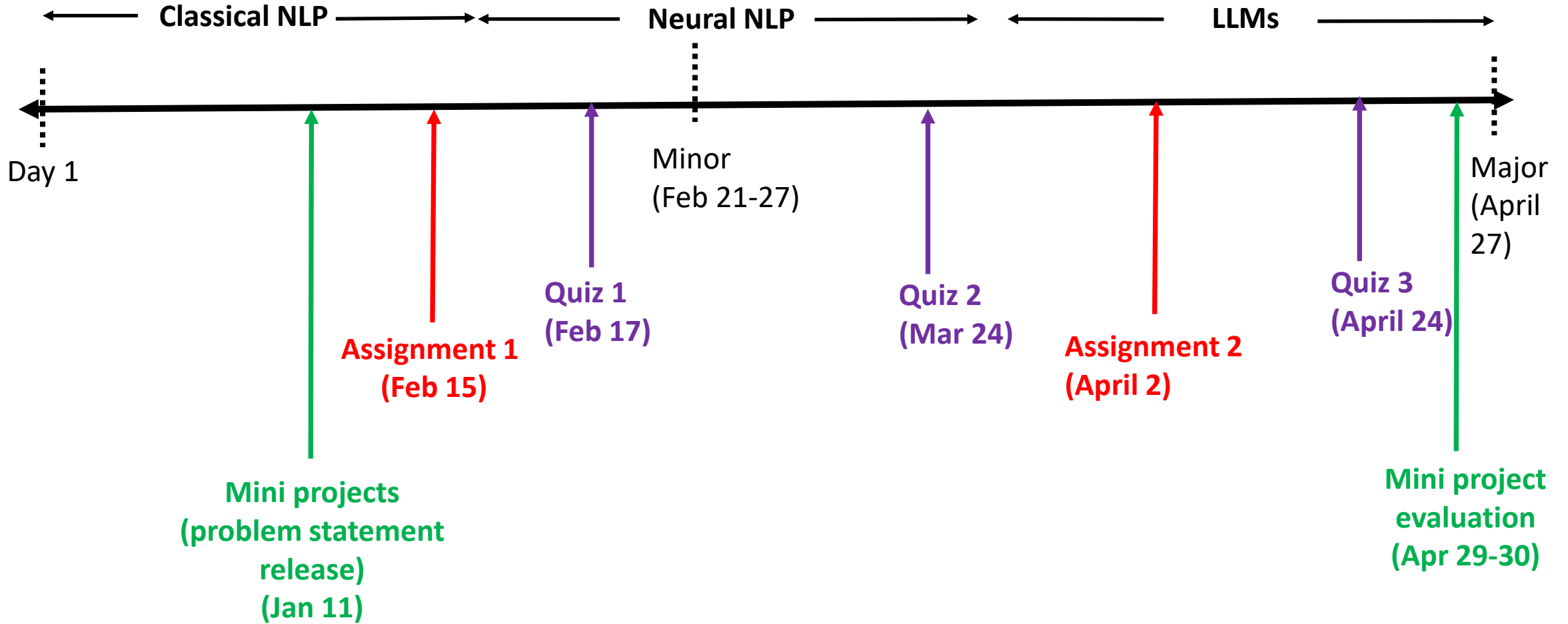
Marks distribution (tentative):

- Minor: 20%
- Major: 30%
- Quiz (3): 15%
- Assignment (2): 15%
- Mini-project: 18% (*group-wise*)
- Paper reading (1)

- **Audit:** Discouraged!
B- (threshold to pass the course)
- **Grading Scheme:** Relative?
- 75% attendance **mandatory**
• If < 75% one grade down

Those who took my LLM course are **discouraged** to take this course.

Timeline



Mini Project (18%)

- A few problem statements, and datasets will be floated (Jan 11, 2025)*
- A leaderboard will be maintained per problem statement
- Each group should consist of **1-2 students?**
- **Best Project Award**
- You need to
 - develop models
 - evaluate your models
 - prepare presentation
 - write tech report

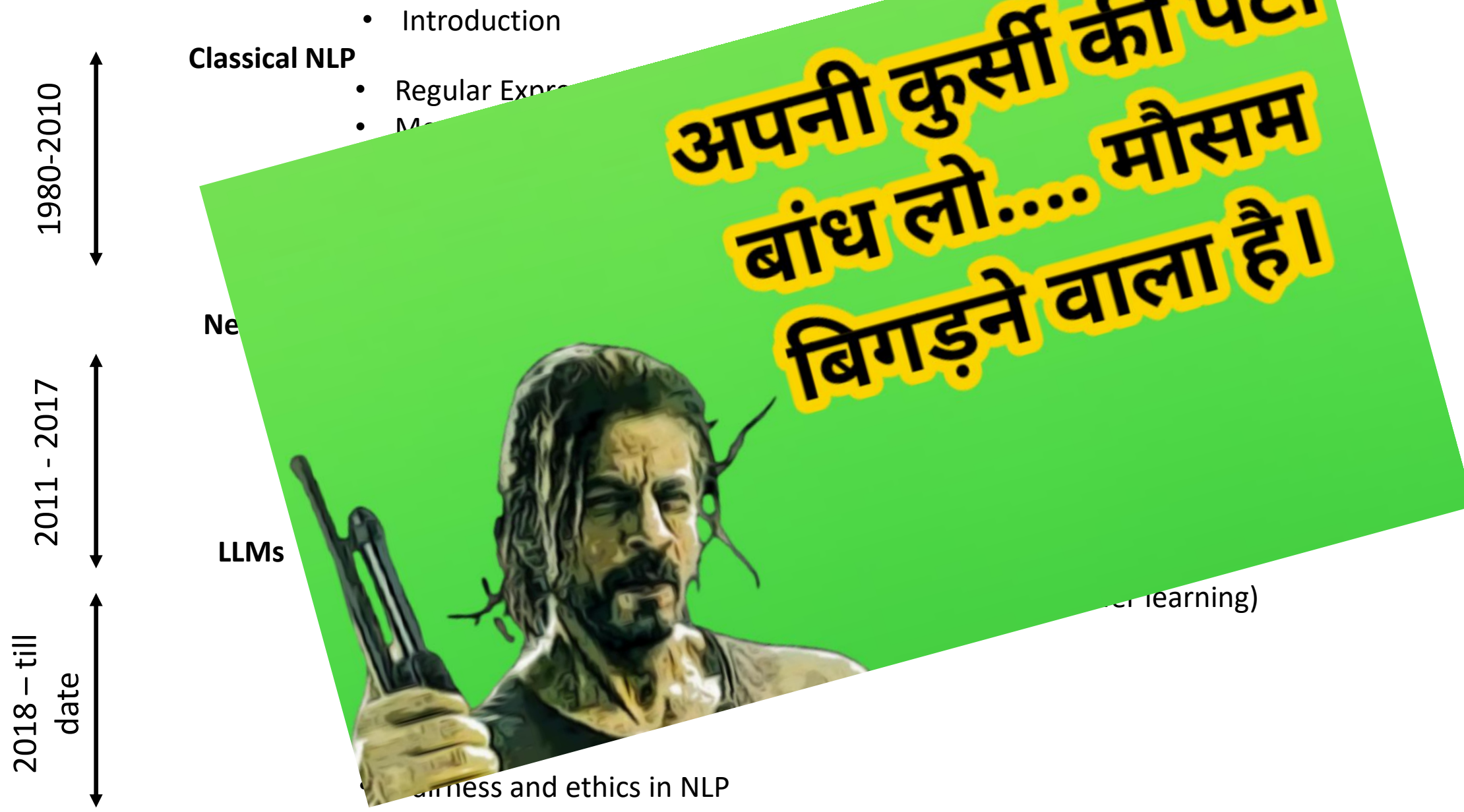
Students are encouraged to publish their projects in good conferences/journals

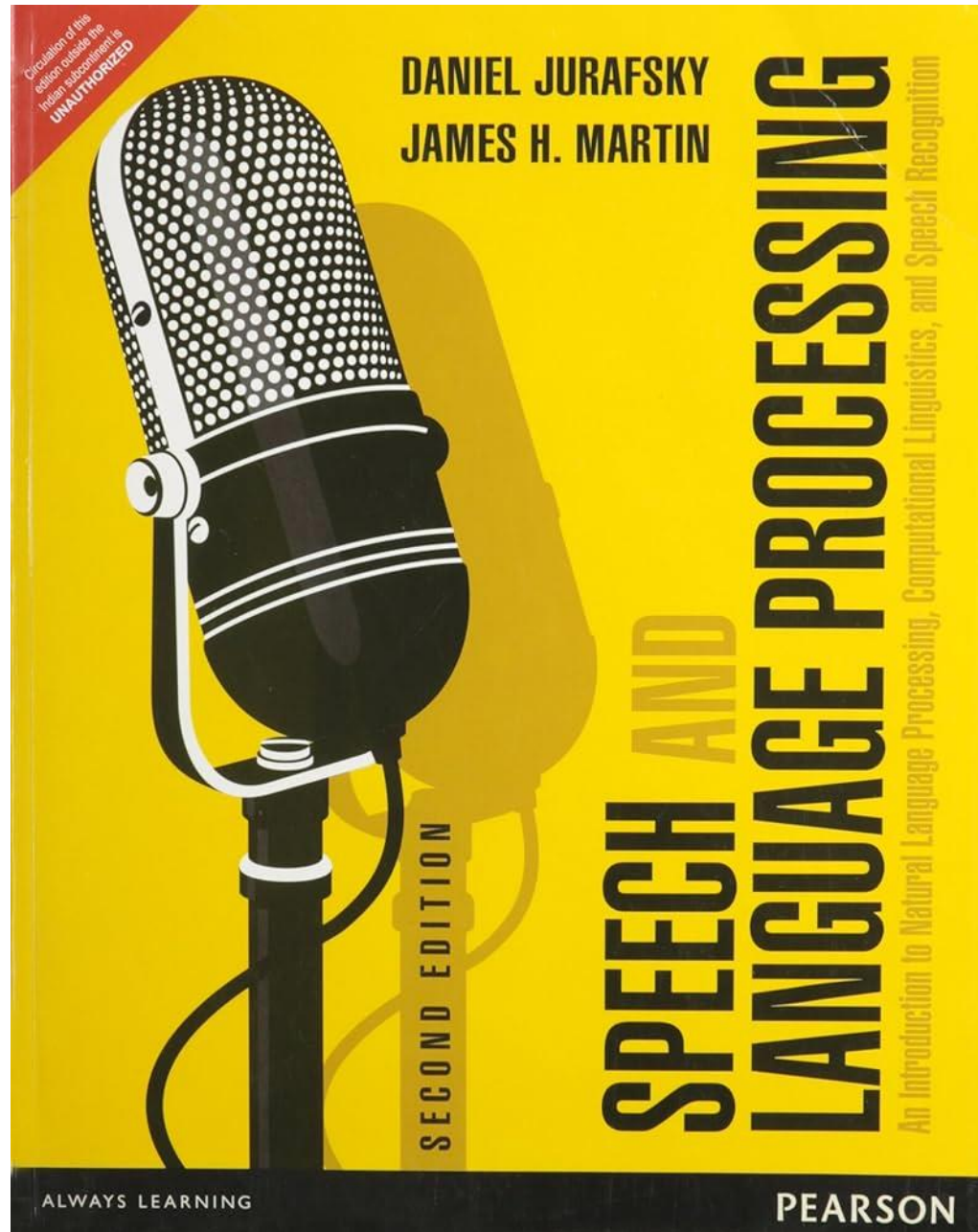
Deliverables:

1. Final project report (**8%**), 8 pages ACL format.
2. Repo of dataset and source code (**5%**)
3. Final project presentation (**5%**)

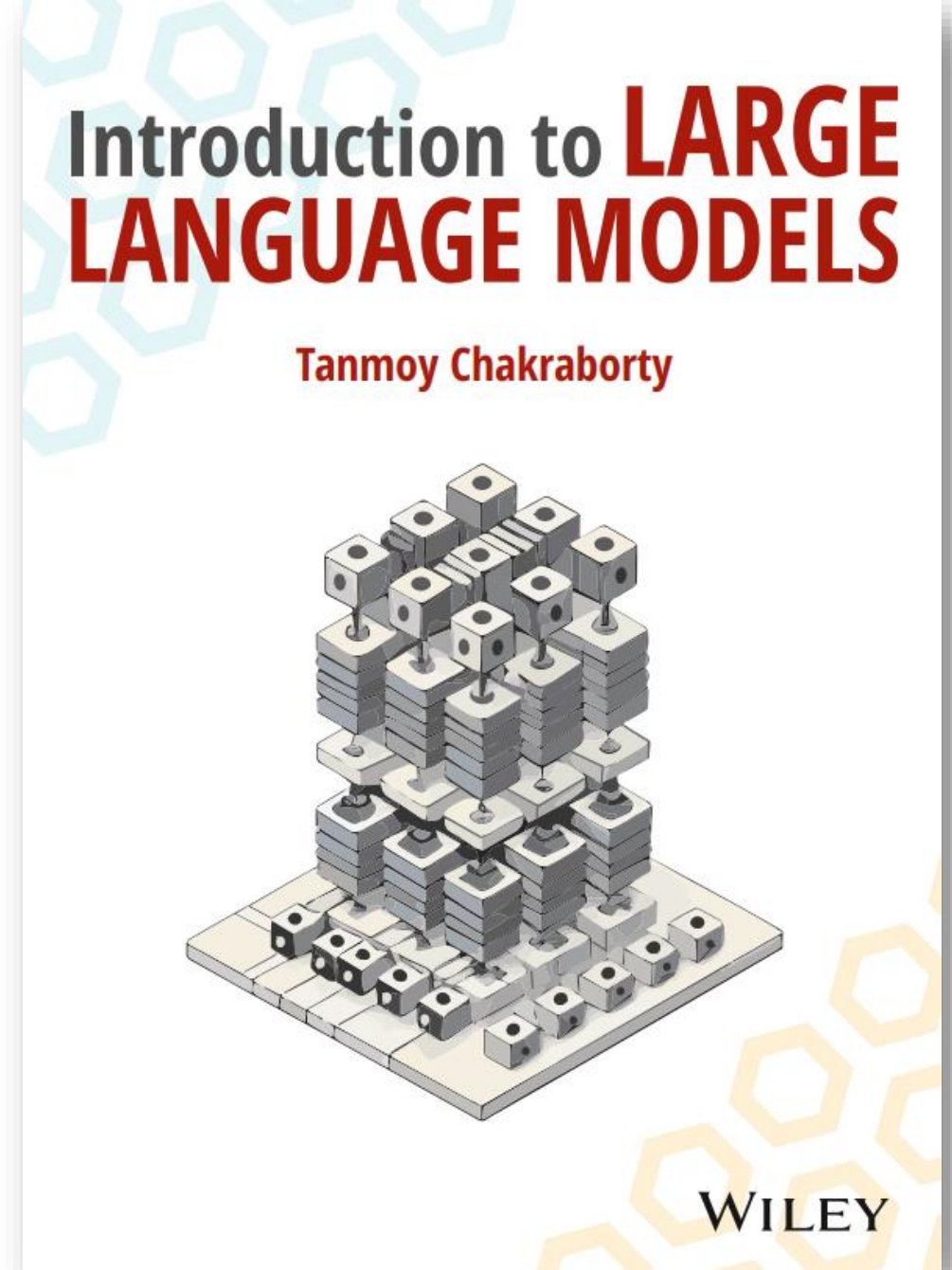
* You are welcome to propose a new idea if you find it fascinating to be qualified for a mini project. Instructor opines!

Content (Tentative)





<https://www.amazon.in/Speech-Language-Processing-Introduction-Computational/dp/9332518416>



<https://www.amazon.in/dp/936386474X/>

Reading and Reference materials

- Journals

- Computational Linguistics, Natural Language Engineering, TACL, KBS, ACM TALLIP,

- Conferences

- ACL, EMNLP, NAACL, COLING, AAI, IJCNLP, ICML, NIPS, WWW, KDD, SIGIR,

Acknowledgment

These slides were adapted from the book

[Introduction to LLMs: https://tanmoychak.com/llmbook/](https://tanmoychak.com/llmbook/)

[SPEECH and LANGUAGE PROCESSING:](#)

[An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition](#)

Advanced NLP, Graham Nuebig <http://www.phontron.com/class/anlp2022/>

Advanced NLP, Mohit Ayyer <https://people.cs.umass.edu/~miyyer/cs685/>

NLP with Deep Learning, Chris Manning, <http://web.stanford.edu/class/cs224n/>

Understanding Large Language Models, Danqi Chen <https://www.cs.princeton.edu/courses/archive/fall22/cos597G/>

and some modifications from presentations found in the WEB by
several scholars including the following

Credits and Acknowledgment

Husni Al-Muhtaseb	Heshaam Feili	Khurshid Ahmad	Martha Palmer
James Martin	Björn Gambäck	Staffan Larsson	julia hirschberg
Jim Martin	Christian Korthals	Robert Wilensky	Elaine Rich
Dan Jurafsky	Thomas G. Dietterich	Feiyu Xu	Christof Monz
Sandiway Fong	Devika Subramanian	Jakub Piskorski	Bonnie J. Dorr
Song young in	Duminda Wijesekera	Rohini Srihari	Nizar Habash
Paula Matuszek	Lee McCluskey	Mark Sanderson	Massimo Poesio
Mary-Angela Papalaskari	David J. Kriegman	Andrew Elks	David Goss-Grubbs
Dick Crouch	Kathleen McKeown	Marc Davis	Thomas K Harris
Tracy Kin	Michael J. Ciaraldi	Ray Larson	John Hutchins
L. Venkata Subramaniam	David Finkel	Jimmy Lin	Alexandros Potamianos
Martin Volk	Min-Yen Kan	Marti Hearst	Mike Rosner
Bruce R. Maxim	Andreas Geyer-Schulz	Andrew McCallum	Latifa Al-Sulaiti
Jan Hajič	Franz J. Kurfess	Nick Kushmerick	Giorgio Satta
Srinath Srinivasa	Tim Finin	Mark Craven	Jerry R. Hobbs
Simeon Ntafos	Nadjet Bouayad	Chia-Hui Chang	Christopher Manning
Paolo Pirjanian	Kathy McCoy	Diana Maynard	Hinrich Schütze
Ricardo Vilalta	Hans Uszkoreit	James Allan	Alexander Gelbukh
Tom Lenaerts	Azadeh Maghsoodi		Gina-Anne Levow
	Md Shad Akhtar		Guitao Gao
	Mohit Ayyer		Qing Ma
	Graham Neubig		Zeynep Altan
	Chris Manning		Edureka
			And many others...

Introduction



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.** **P 17**

Indian cricket board to release **₹1cr in financial assistance to ex-cricket 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 6**

> **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**

Gang guns down Trinamool man at Bengal dhaba

Dapi 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-

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

Getty Images/USA

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-presz/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.Kalital
@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

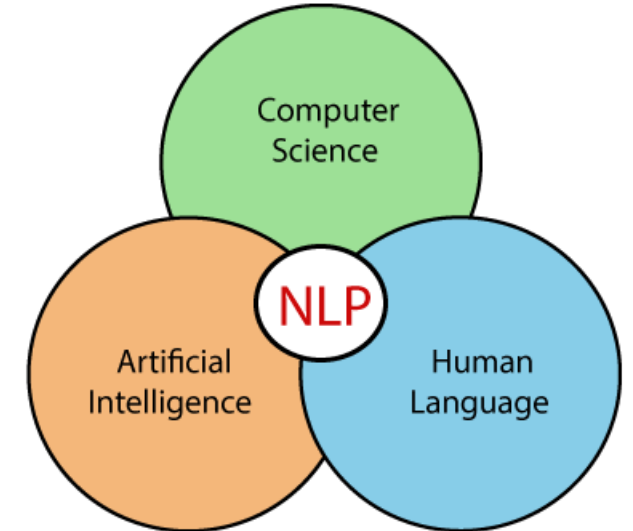
Is this a grammatically correct English sentence?

Buffalo buffalo Buffalo buffalo buffalo buffalo Buffalo buffalo

Natural Language Processing

- **What is a Natural Language?**

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



- **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.

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.

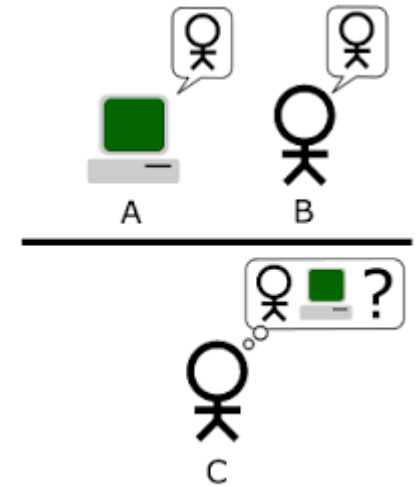
<https://indianexpress.com/article/india/more-than-19500-mother-tongues-spoken-in-india-census-5241056/>



Natural Language Processing

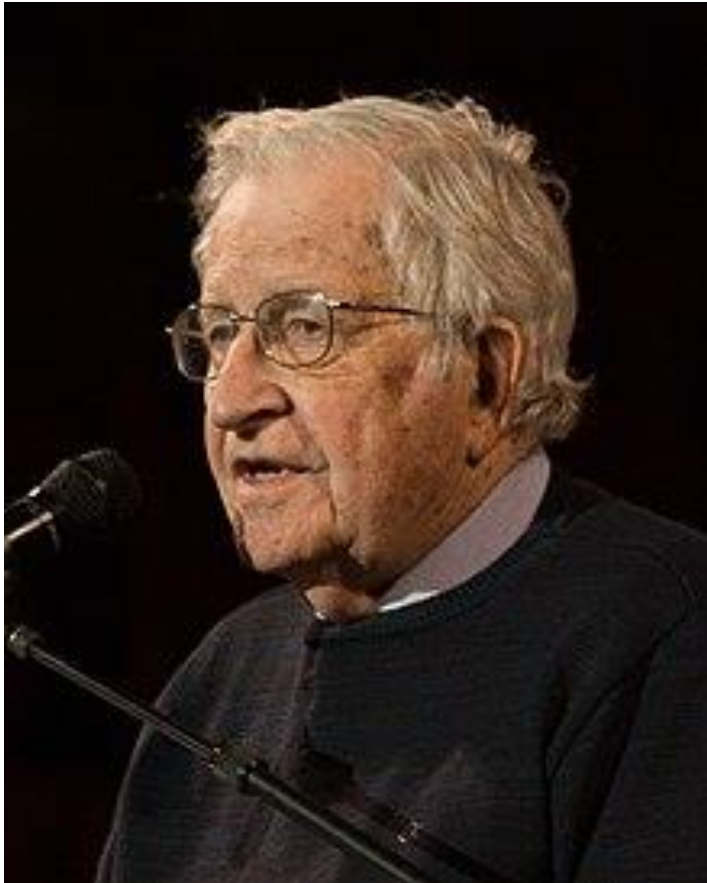


- Setup
 - Two rooms, two humans, and a computer.
 - Room 1: One human C
 - Room 2: One computer (A) and one human (B)
 - A response generated from room 2 (either by A or B)
 - C has to figure out the source of the response
 - If C is successful → "A" failed the turing test
 - Else, → "A" passed the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



"Computing Machinery and Intelligence" which 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 the [University of Arizona](#) and an [institute professor](#) emeritus at the [MIT](#)

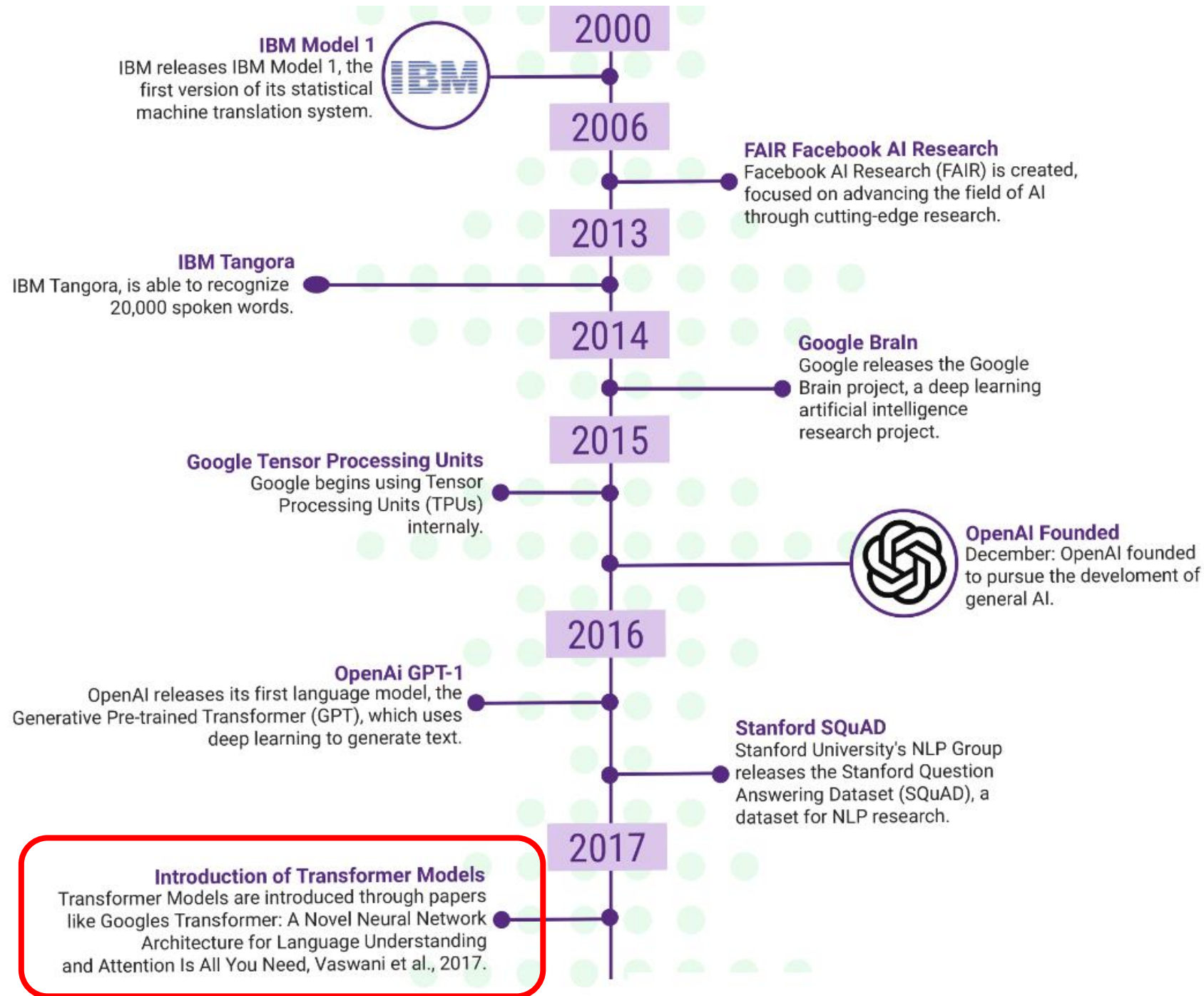
Natural Language Processing

Software	Year	Creator	Description
Georgetown experiment	1954	Georgetown University and IBM	involved fully automatic translation of more than sixty Russian sentences into English.
STUDENT	1964	Daniel Bobrow	could solve high school algebra word problems. ^[6]
ELIZA	1964	Joseph Weizenbaum	a simulation of a Rogerian psychiatrist , rephrasing her response with a few grammar rules. ^[7]
SHRDLU	1970	Terry Winograd	a natural language system working in restricted "blocks worlds" with restricted vocabularies, worked extremely well
PARRY	1972	Kenneth Colby	A chatbot
KL-ONE	1974	Sondheimer et al.	a knowledge representation system in the tradition of semantic networks and frames; it is a frame language .
MARGIE	1975	Roger Schank	
TaleSpin (software)	1976	Meehan	
QUALM		Lehnert	
LIFER/LADDER	1978	Hendrix	a natural language interface to a database of information about US Navy ships.
SAM (software)	1978	Cullingford	
PAM (software)	1978	Robert Wilensky	
Politics (software)	1979	Carbonell	
Plot Units (software)	1981	Lehnert	
Jabberwacky	1982	Rollo Carpenter	chatbot with stated aim to "simulate natural human chat in an interesting, entertaining and humorous manner".
MUMBLE (software)	1982	McDonald	
Racter	1983	William Chamberlain and Thomas Etter	chatbot that generated English language prose at random.
MOPTRANS ^[8]	1984	Linzen	
KODIAK (software)	1984	Wilensky	
Absity (software)	1987	Hirst	
Dr. Sbaits	1991	Creative Labs	
Watson (artificial intelligence software)	2006	IBM	A question answering system that won the Jeopardy! contest, defeating the best human players in February 2011.
Siri	2011	Apple	A virtual assistant developed by Apple.
Cortana	2014	Microsoft	A virtual assistant developed by Microsoft.
Amazon Alexa	2014	Amazon	A virtual assistant developed by Amazon.
Google Assistant	2016	Google	A virtual assistant developed by Google.

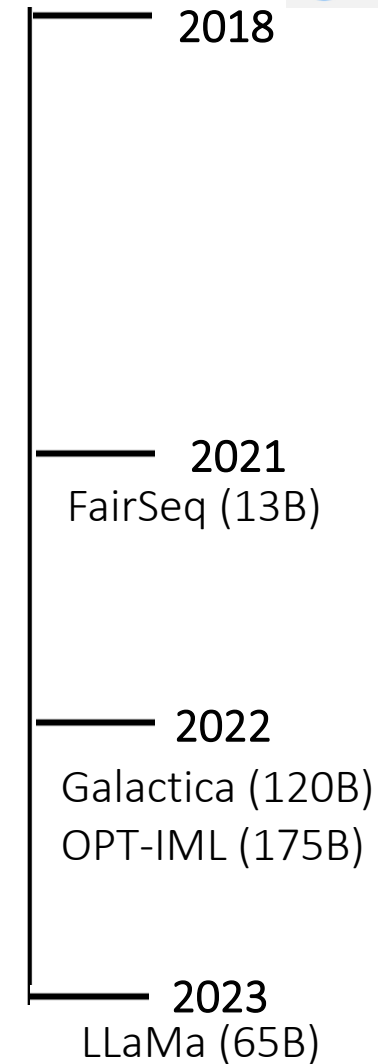
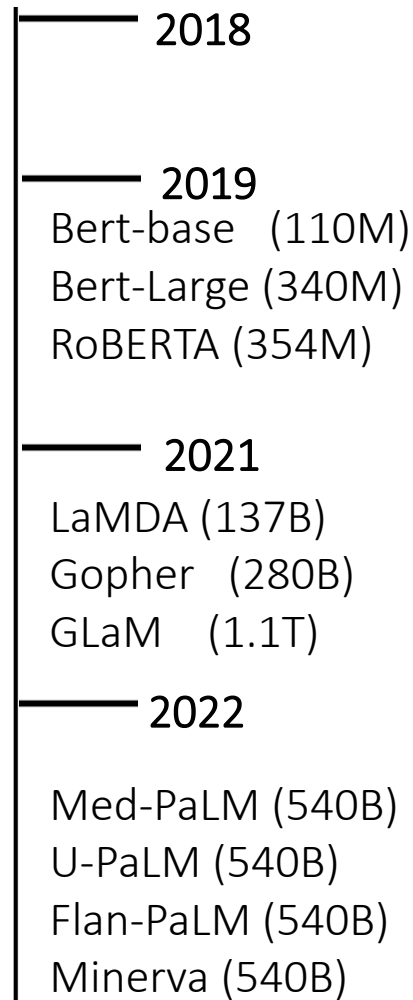
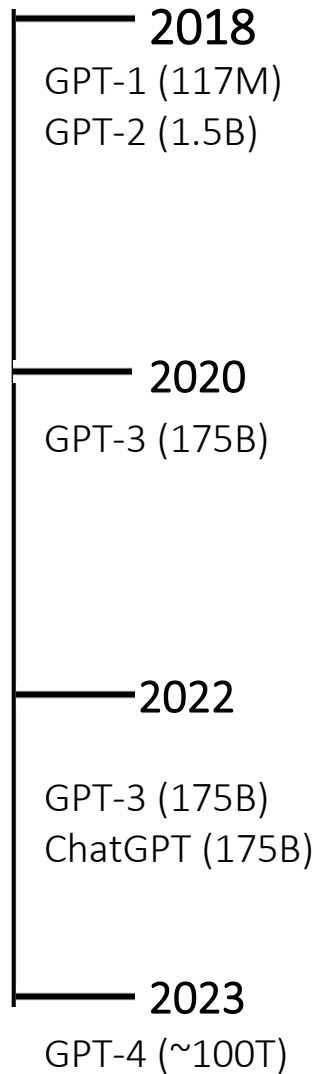
Natural Language Processing

Software	Year	Creator	Description
Georgetown experiment	1954	Georgetown University and IBM	involved fully automatic translation of more than sixty Russian sentences into English.
STUDENT	1964	Daniel Bobrow	could solve high school algebra word problems. ^[6]
ELIZA	1964	Joseph Weizenbaum	a simulation of a Rogean psychiatrist, rephrasing her response with a few grammar rules. ^[7]
SHRDLU	1970	Terry Winograd	a natural language system working in restricted "blocks worlds" with restricted vocabularies, worked extremely well
PARRY	1972	Kenneth Colby	A chatterbot
KL-ONE	1974	Sondheimer et al.	a knowledge representation system in the tradition of semantic networks and frames; it is a frame language .
MARGIE	1975	Roger Schank	
TaleSpin (software)	1976	Meehan	
QUALM		Lehnert	
LIFER/LADDER	1978	Hendrix	a natural language interface to a database of information about US Navy ships.
SAM (software)	1978	Cullingford	
PAM (software)	1978	Robert Wilensky	
Politics (software)	1979	Carbonell	
Plot Units (software)	1981	Lehnert	
Jabberwacky	1982	Rollo Carpenter	chatterbot with stated aim to "simulate natural human chat in an interesting, entertaining and humorous manner".
MUMBLE (software)	1982	McDonald	
Racter	1983	William Schabberlain and Thomas Etter	chatterbot that generated English language prose at random.
MOPTRANS ^[8]	1984	McNien	
KODIAK (software)	1984	Wilensky	
Absity (software)	1987	Hirst	
Dr. Sbaitso	1991	Creative Labs	
Watson (artificial intelligence software)	2006	IBM	A question answering system that won the Jeopardy! contest, defeating the best human players in February 2011.
Siri	2011	Apple	A virtual assistant developed by Apple.
Cortana	2014	Microsoft	A virtual assistant developed by Microsoft.
Amazon Alexa	2014	Amazon	A virtual assistant developed by Amazon.
Google Assistant	2016	Google	A virtual assistant developed by Google.

Timeline History of Large Language Models



Post-Transformer Landscape



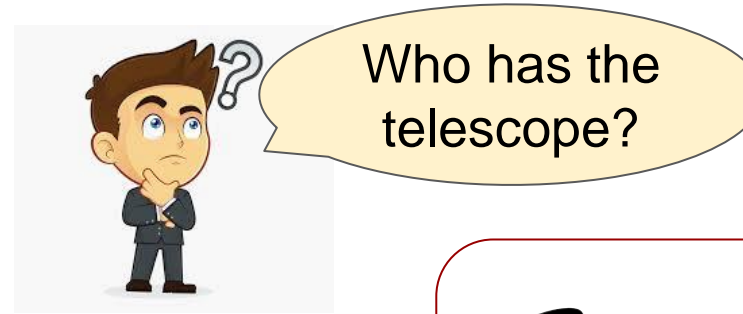
Ambiguity

- Is ambiguity present in language only?
 - No, ambiguity is prevalent in every dimension!

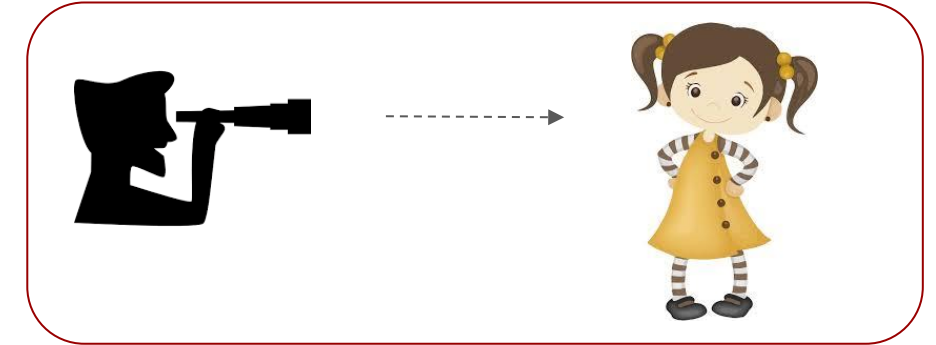
Duck or Rabbit?



Ambiguity in language



- I saw a girl with a telescope.



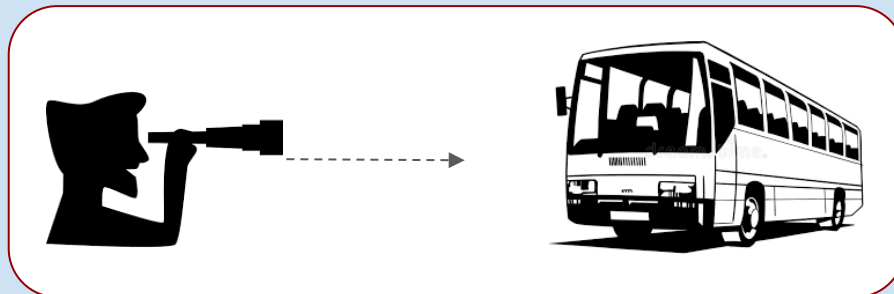
OR



- I saw a girl with a bicycle.



- I saw a bus with a telescope.



No ambiguity!

Ambiguity in language



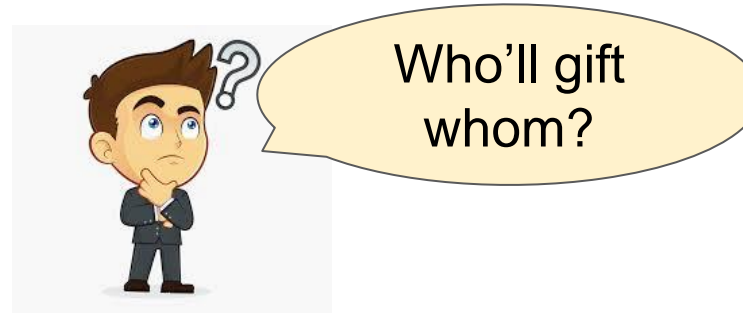
- I saw a girl with a telescope.
- Mary had a little lamb.



OR



Ambiguity in language



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

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 padegei.
- **Public demand changes**



OR



- (a) Public demand changes, but does anybody listen to them?
(b) Public demand changes, and we companies have to adapt to such changes.

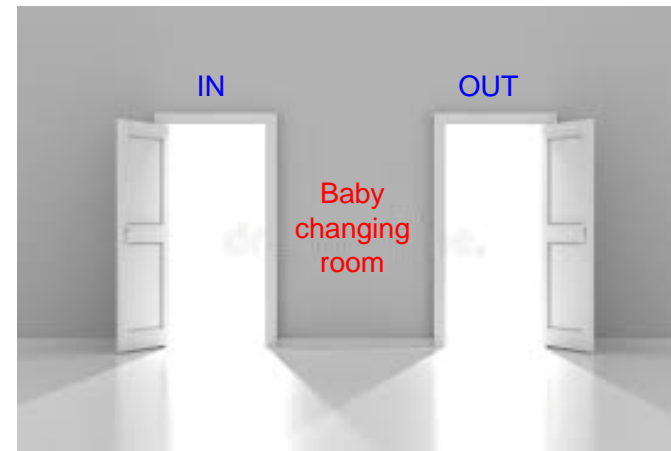
Ambiguity in language



- I saw a girl with a telescope.
- Mary had a little lamb.
- Mujhe aapko mithai khilani padegei.
- Public demand changes
- Baby changing room



OR



Ambiguity in language

- I saw a girl with a telescope.
- Mary had a little lamb.
- Mujhe aapko mithai khilani padegi.
- Public demand changes
- Baby changing room
- I ate rice with spoon.
- I ate rice with curd.
- I ate rice with Rahul.



Similar surface
structures but
different
interpretations!



Ambiguity and Punctuations!

Let's eat Grandma!



Let's eat, Grandma!



A woman without her man is nothing

A woman, without her man, is nothing.

A woman: without her, man is nothing.

Punctuation is powerful.



Ambiguity makes NLP hard

Surface form has multiple interpretations

- **Syntactic Ambiguity**

- Violinist Linked to JAL Crash Blossoms => main verb?

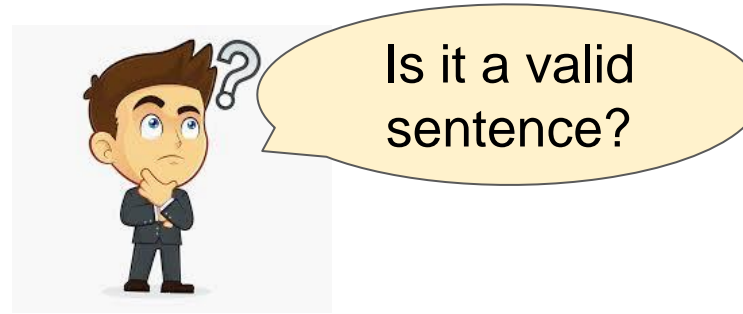
a violinist who is linked to a crash is **blossoming in her career**

Etymology [\[edit\]](#)

From a headline "Violinist linked to JAL crash blossoms". The author's intended interpretation is that the violinist who blossoms was linked to a plane crash (by her father having been on the plane). However, the sentence can also be interpreted to mean that the violinist was linked to something called a "crash blossom".

the study of the origin of words and the way in which their meanings have changed throughout history.

What about this?

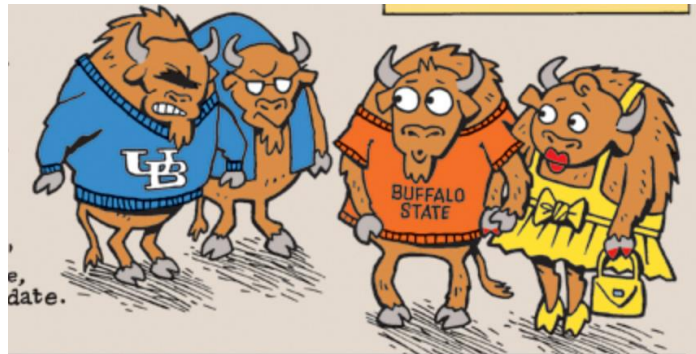


Buffalo buffalo Buffalo buffalo buffalo buffalo Buffalo buffalo

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



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 York Railroad
the New York Central 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 ...

that's what makes it fun!

Components of NLP



**Natural Language
Understanding**



**Natural Language
Generation**

NLP layers

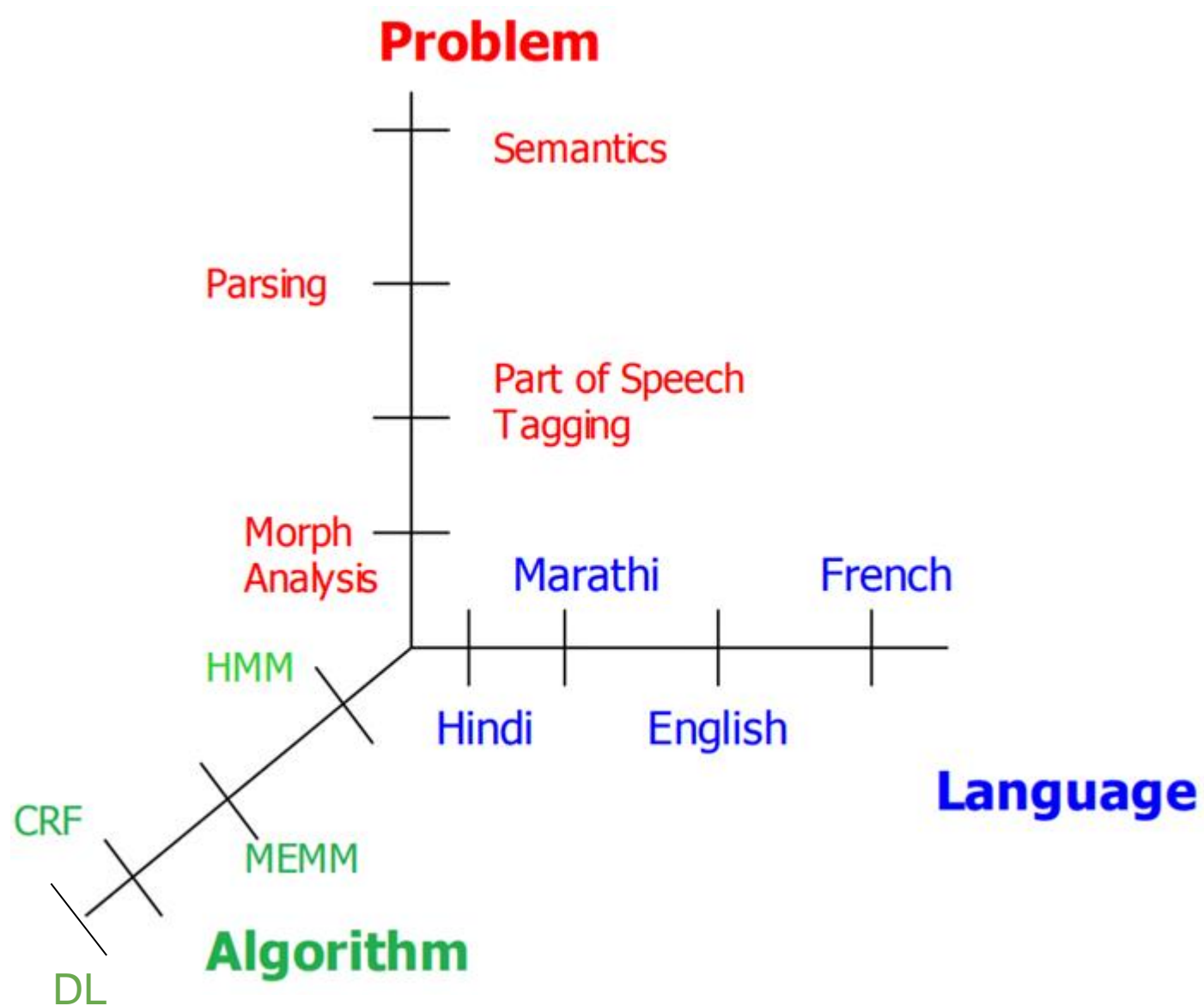
- Understanding the semantics is a non-trivial task.
- Needs to performs a series of incremental tasks to achieve this.
- NLP happens in layers

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



Increasing
Complexity Of
Processing

NLP trinity



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.
 - 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

- 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
 - E.g.:
 - “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.
- Indian, 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।	

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

- PoS disambiguation
 - A word can belong to different grammatical classes.

He	went	to	the	<i>park</i>	in	a	car	.
PRP	VBD	TO	DT	<i>NN</i>	IN	DT	NN	.



They	went	to	<i>park</i>	the	car	in	the	shed	.
PRP	VBD	TO	<i>VB</i>	DT	NN	IN	DT	NN	.

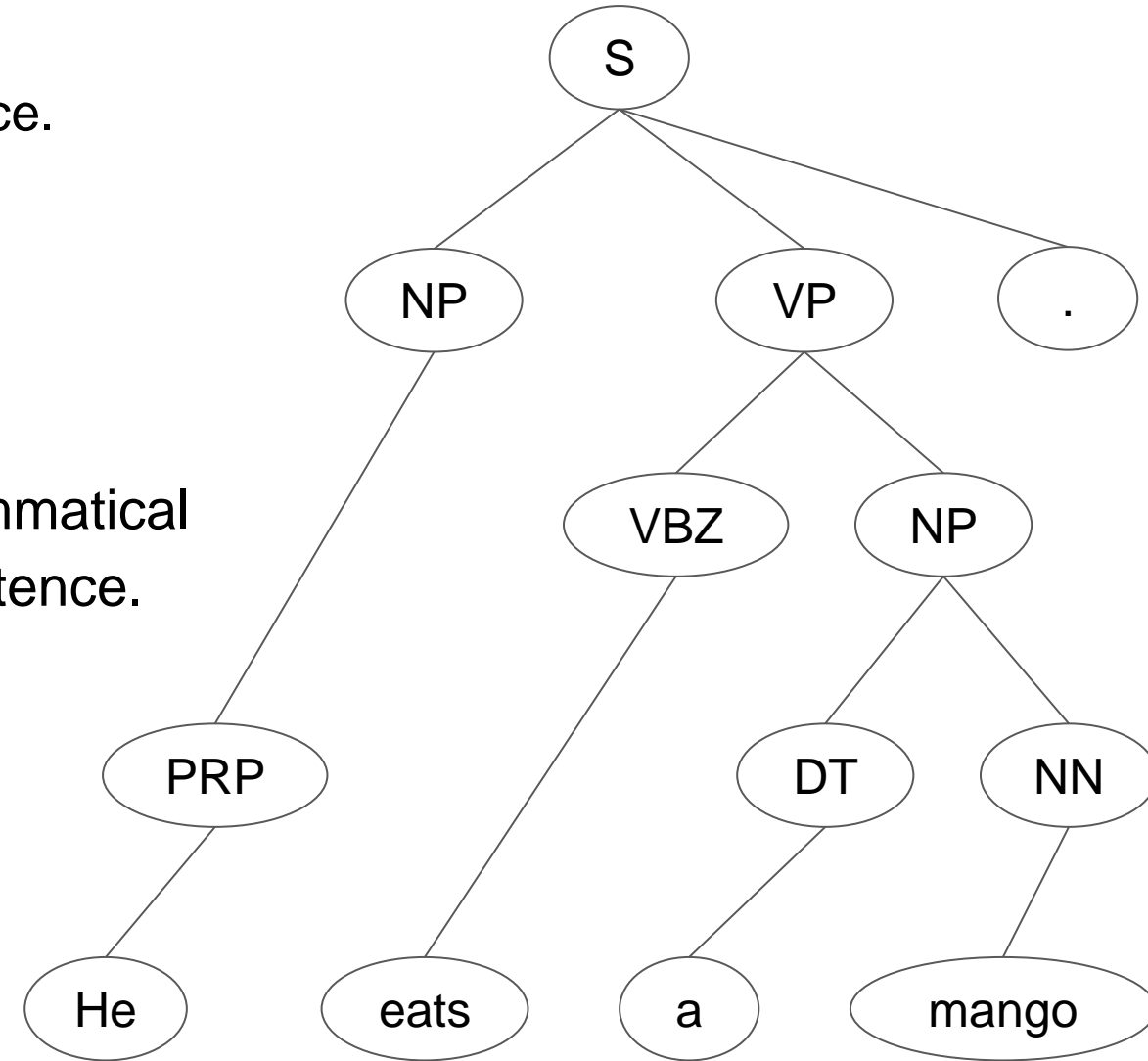
- 45 tags in Penn Treebank tagset
- 146 tags in C7

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]

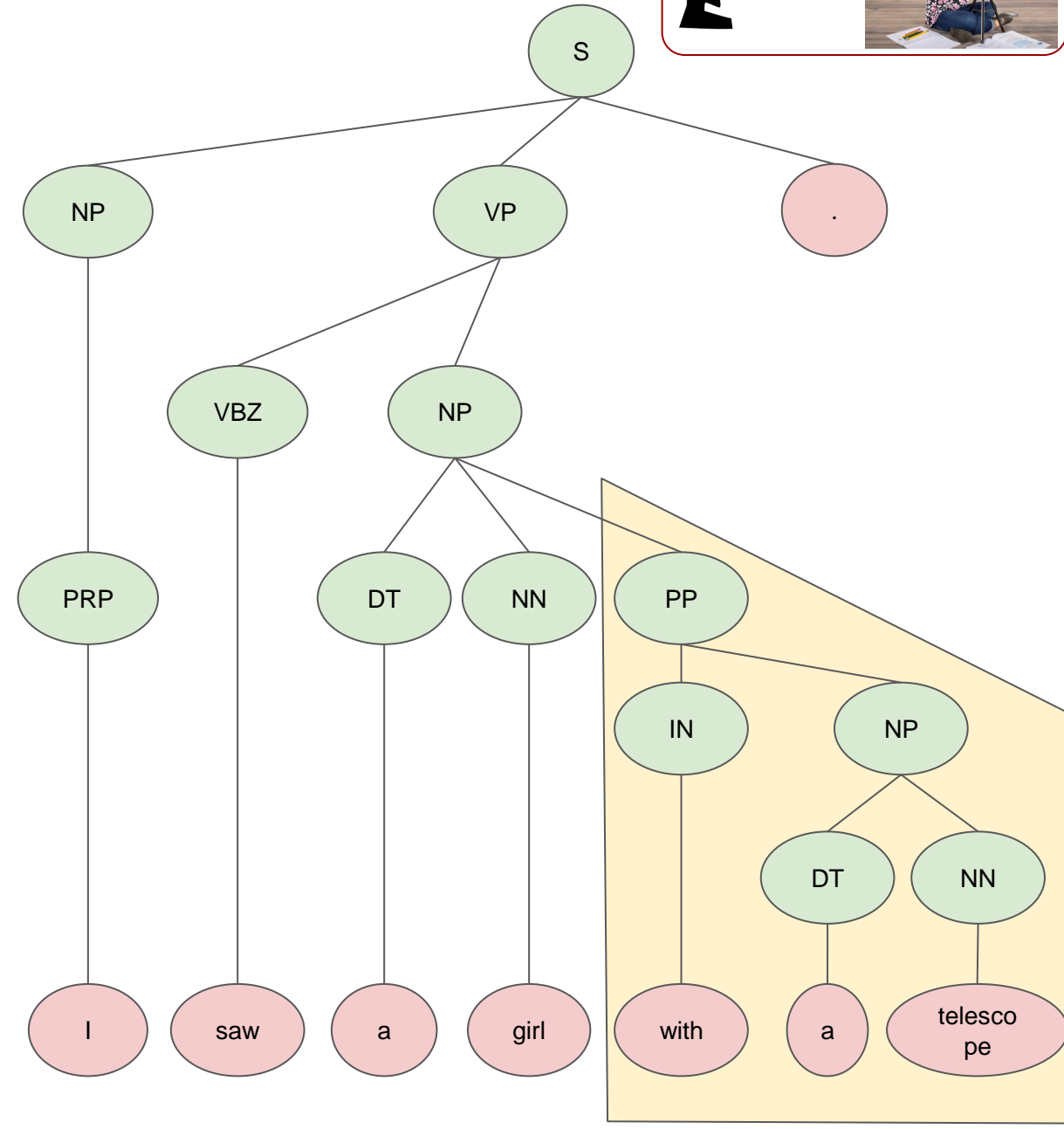
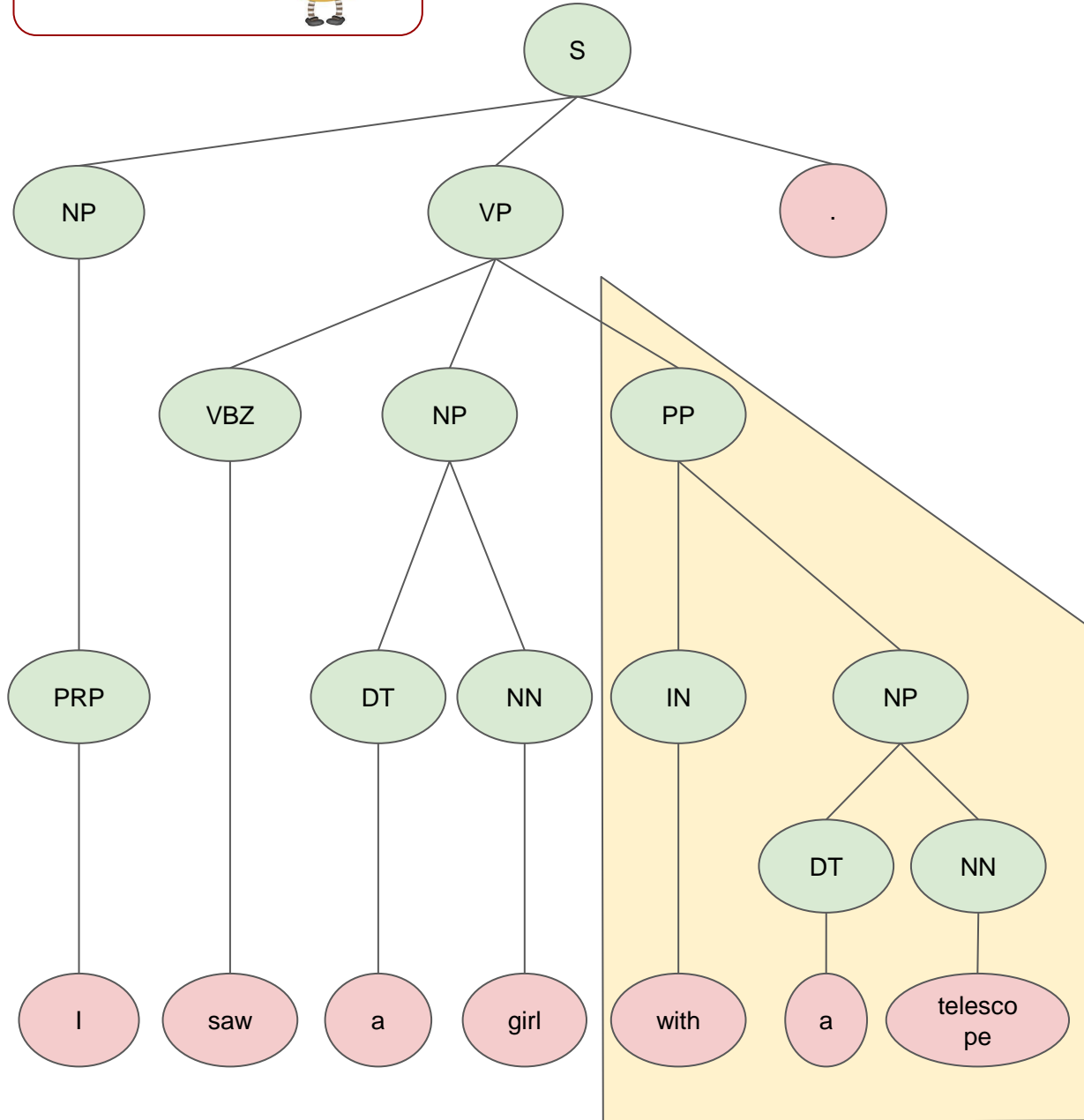
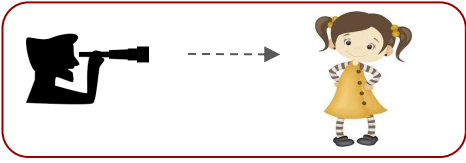
Syntax Processing

- Validate the grammatical structure of the sentence.
- 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.



Parse Tree

Syntactic Ambiguity



Tasks we want to solve in NLP

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

- 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

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

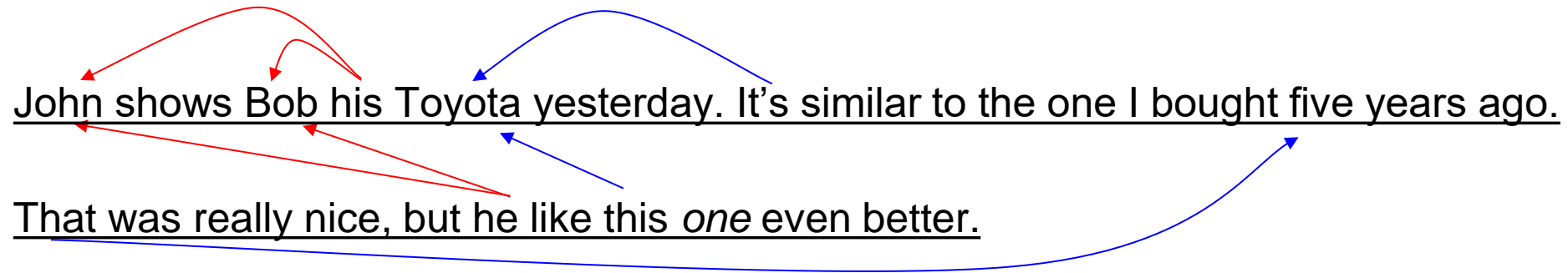
- Processing of sequence of sentences.

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?

Coreference Resolution

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

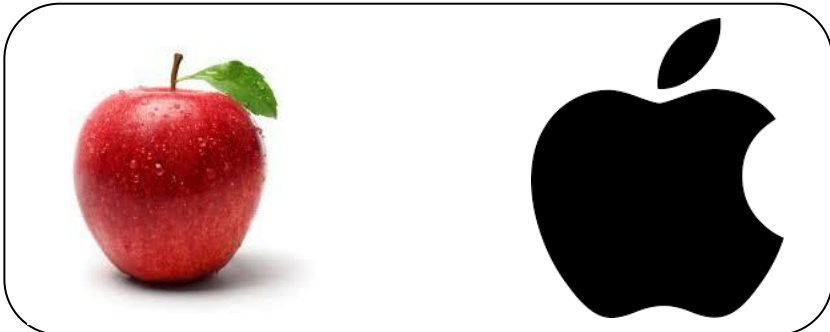


Information Extraction

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

- 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 hrs long. ⇒ Neutral?

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.

ENGLISH - DETECTED SOMALI ENG ▼ ↔ HINDI SOMALI ENGLISH ▼

I saw a girl with telescope. × मैंने दूरबीन से एक लड़की को देखा।

English ▼ ↔ Hindi ▼

She is a doctor × वह एक डॉक्टर है
vah ek doktor hai

Hindi ▼ ↔ English ▼

वह एक डॉक्टर है × He is a doctor



Source: Google Translate

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.

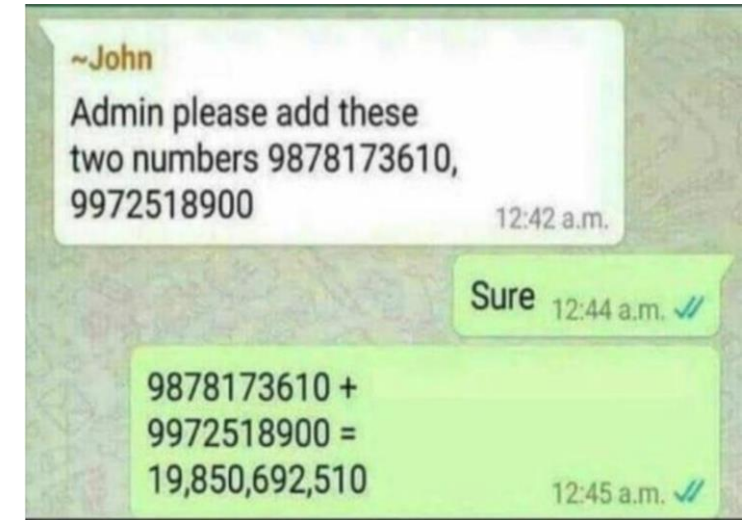
Question Answering

- 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

- Conversation of two or more parties.

C₁: ...I need to travel in May.
A₁: And, what day in May did you want to travel?
C₂: OK uh I need to be there for a meeting that's from the 12th to the 15th.
A₂: And you're flying into what city?
C₃: Seattle.
A₃: And what time would you like to leave Pittsburgh?
C₄: Uh hmm I don't think there's many options for non-stop.
A₄: Right. There's three non-stops today.
C₅: What are they?
A₅: 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.
C₆: OK I'll take the 5ish flight on the night before on the 11th.
A₆: On the 11th? OK. Departing at 5:55pm arrives Seattle at 8pm, U.S. Air flight 115.
C₇: 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.

Why Study NLP?

- To get a job in industry
 - e.g., many current job listings are CL jobs
 - Google Inc.
 - Amazon Inc.
 - Facebook Inc.
 - Flipkart Inc., etc.
- To get a job in academia
 - As a computational linguist
 - computational literacy and an understanding of computational methods will become critical in the next decade.