



Introduction to NLP (ELL 881)

Special Topics in Computers 2



Neuro-linguistic programming

Introduction to NLP (ELL 881)
Special Topics in Computers 2



Neuro-linguistic
programming

Introduction to NLP (ELL 881)

Special Topics in Computers 2



Non-Linear Programming

Introduction to NLP (ELL 881)

Special Topics in Computers 2



Non-Linear Programming

Introduction to NLP (ELL 881)
Special Topics in Computers 2



Natural Language Processing

Introduction to NLP (ELL 881)
Special Topics in Computers 2

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 (tanmoychak.com)
(NLP, Social Media, Graph Neural Networks)
tanchak@iitd.ac.in
- Guest Lecture: TBD
- Course page: <https://sites.google.com/view/ell881-iitd/home>
- Piazza: <https://piazza.com/iitd.ac.in/spring2023/ell881>
- TAs:
 - Kshitij Alwadhi (Kshitij.Alwadhi.ee119@ee.iitd.ac.in)
 - Gurusha Juneja (ee1190480@ee.iitd.ac.in)
- 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

Reading and Reference materials

- Books

- Speech and Language Processing, Dan Jurafsky and James H. Martin
<https://web.stanford.edu/~jurafsky/slp3/>
- Foundations of Statistical Natural Language Processing, Chris Manning and Hinrich Schütze
- Natural Language Processing, Jacob Eisenstein
<https://github.com/jacobeisenstein/gt-nlp-class/blob/master/notes/eisenstein-nlp-notes.pdf>
- A Primer on Neural Network Models for Natural Language Processing, Yoav Goldberg
<http://u.cs.biu.ac.il/~yogo/nnlp.pdf>

- Journals

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

- Conferences

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

Research papers repository

<https://aclanthology.org/>

Research papers repository

<https://arxiv.org/list/cs.CL/recent>

arXiv.org > cs > cs.CL

Computation and Language

Authors and titles for recent submissions

- [Wed, 19 Aug 2020](#)
- [Tue, 18 Aug 2020](#)
- [Mon, 17 Aug 2020](#)
- [Fri, 14 Aug 2020](#)
- [Thu, 13 Aug 2020](#)

[total of 84 entries: 1-25 | 26-50 | 51-75 | 76-84]
[showing 25 entries per page: [fewer](#) | [more](#) | [all](#)]

Wed, 19 Aug 2020

[1] [arXiv:2008.07905 \[pdf, other\]](#)

Glancing Transformer for Non-Autoregressive Neural Machine Translation

[Lihua Qian](#), [Hao Zhou](#), [Yu Bao](#), [Mingxuan Wang](#), [Lin Qiu](#), [Weinan Zhang](#), [Yong Yu](#), [Lei Li](#)

Comments: 11 pages, 3 figures, 4 tables

Subjects: Computation and Language (cs.CL)

[2] [arXiv:2008.07880 \[pdf, other\]](#)

COVID-SEE: Scientific Evidence Explorer for COVID-19 Related Research

[Karin Verspoor](#), [Simon Šuster](#), [Yulia Otmakhova](#), [Shevon Mendis](#), [Zenan Zhai](#), [Biaoyan Fang](#), [Jey Han Lau](#), [Timothy Bal](#)

Comments: COVID-SEE is available at [this http URL](#)

Subjects: Computation and Language (cs.CL); Information Retrieval (cs.IR)

[3] [arXiv:2008.07772 \[pdf, other\]](#)

Very Deep Transformers for Neural Machine Translation

[Xiaodong Liu](#), [Kevin Duh](#), [Liyuan Liu](#), [Jianfeng Gao](#)

Comments: 6 pages, 3 figures and 3 tables

Subjects: Computation and Language (cs.CL)

[4] [arXiv:2008.07723 \[pdf, other\]](#)

NASE: Learning Knowledge Graph Embedding for Link Prediction via Neural Architecture Search

[Xiaoyu Kou](#), [Bingfeng Luo](#), [Huang Hu](#), [Yan Zhang](#)

Comments: Accepted by CIKM 2020, short paper

Subjects: Computation and Language (cs.CL)

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.
- Instructor/TAs may cover DL-related prerequisites

Course Directives

- **Class Time:** Mon & Thu, 2 pm – 3:30 pm
- **Office Hour:** Mon 5-6 pm
- **Room:** LH-519



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 1: 10%
- Minor 2: 10%
- Major: 20%
- Quiz (3): 15%
- Assignment (2): 20%
- Mini-project: 20% (**group-wise**)
- Paper reading (1): 5% (**group-wise?**)

- **Audit:** **Discouraged!**
B- (threshold to pass the course)
- **Grading Scheme:** Relative?
- 75% attendance **mandatory** (Timble)
 - If you want to deregister, please do it ASAP
 - Please allow others to register
 - Registration limit (80) may not be increased

Mini Project (20%)

- A few problem statements, and datasets will be floated (in Jan 2023)*
- A leaderboard will be maintained per problem statement
- Each group should consist of **1-3 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:

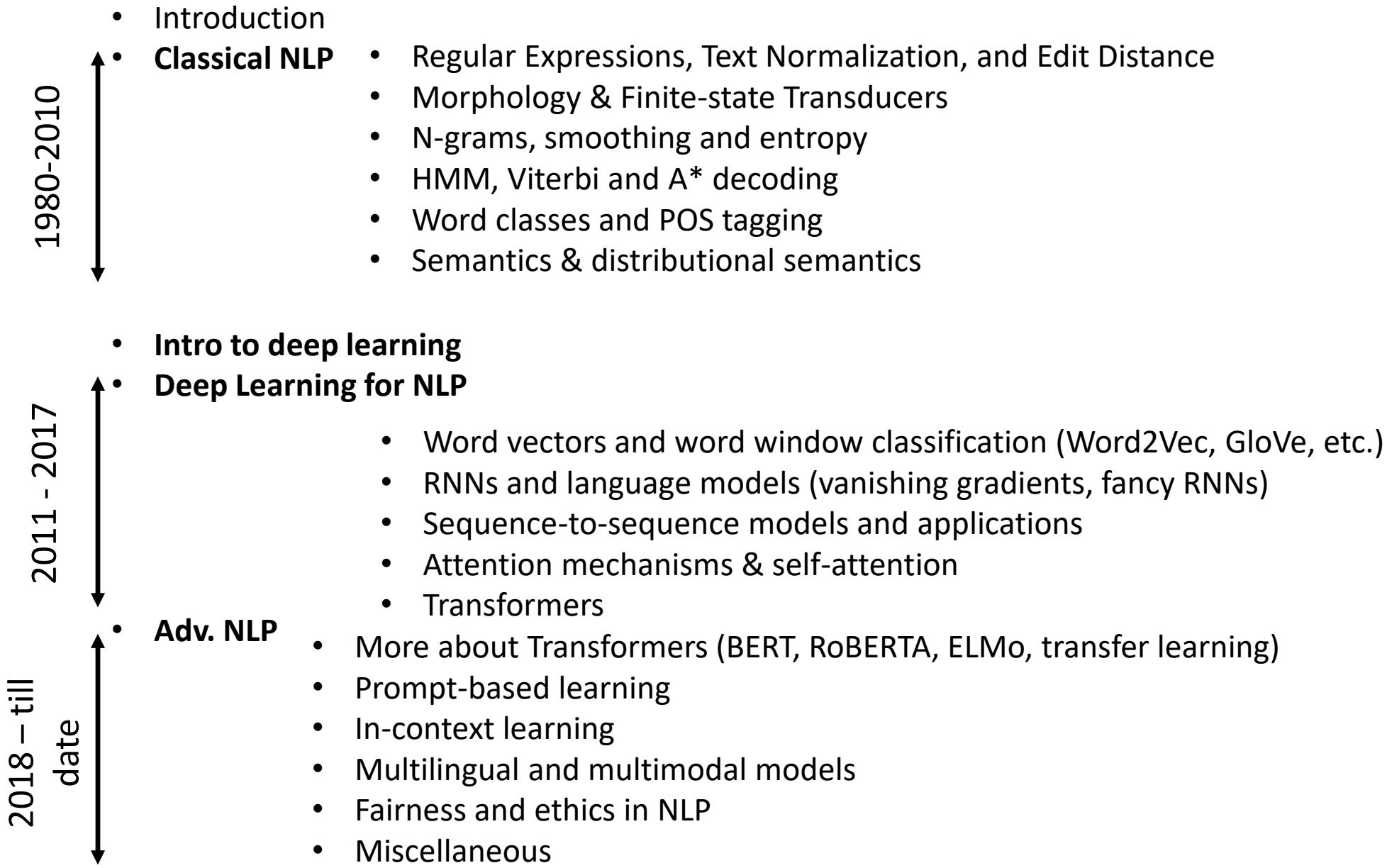
1. Final project report (**8%**), 8 pages ACL format. Need to arxiv
2. Repo of dataset and source code (**2%**)
3. Final project presentation (**5%**)
4. Performance on leaderboard (**5%**)

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

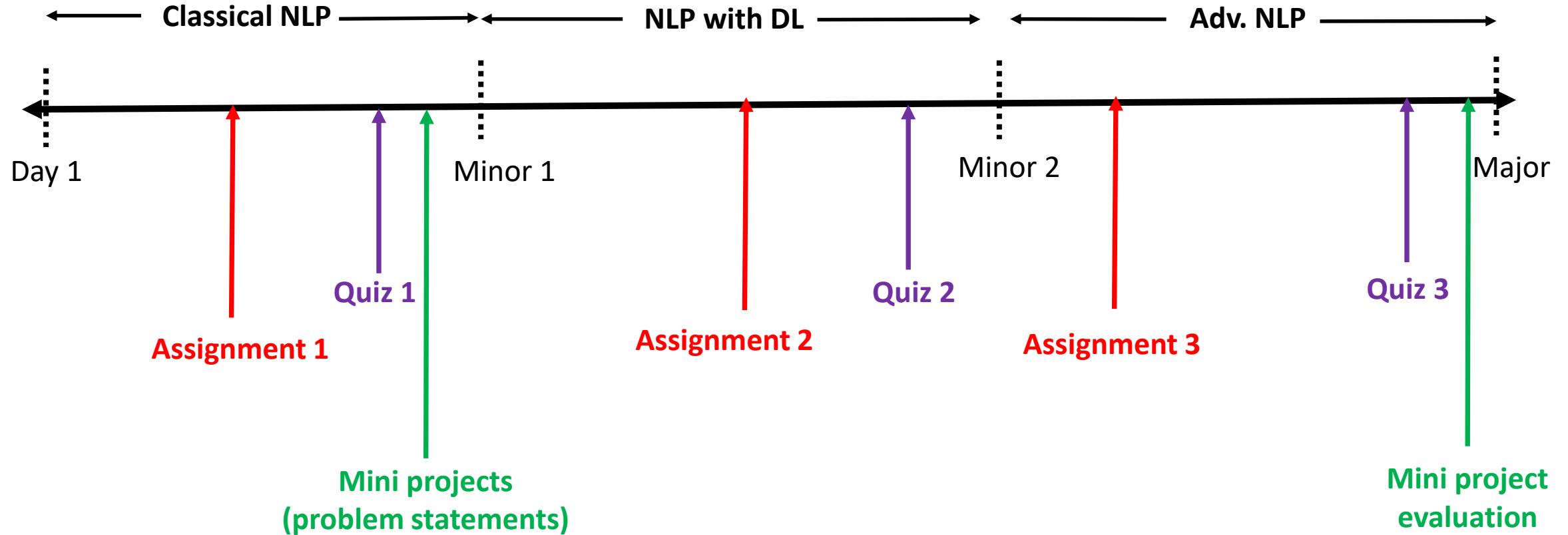
List of Projects

- TBD

Content (Tentative)



Timeline



Two Assignments: $\text{Max}(\text{Assignment 1}, \text{Assignment 2}) + \text{Assignment 3}$

Acknowledgment

These slides were adapted from the book

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 Hajíč	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 Maghsoudi		Gina-Anne Levow
	Md Shad Akhtar		Guitao Gao
	Mohit Ayyer		Qing Ma
	Graham Neubig		Zeynep Altan
	Chris Manning		Edureka
			And many others...

Introduction

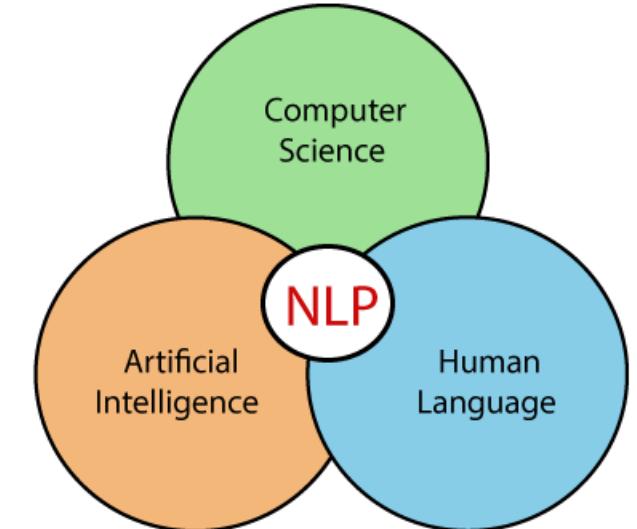
Is this a grammatically correct English sentence?

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.

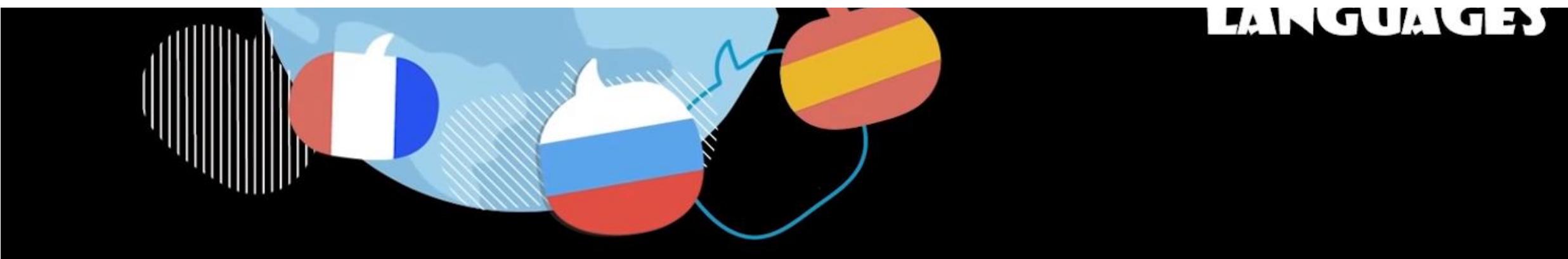
The Human Language

[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.

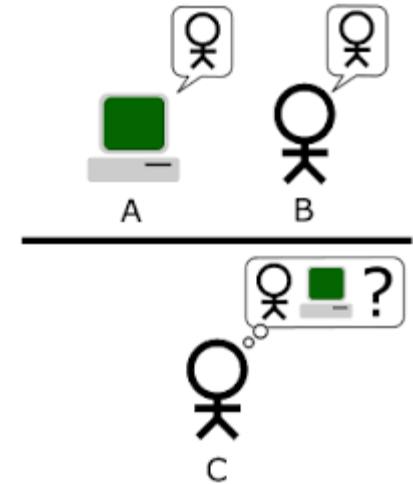
<https://indianexpress.com/article/india/more-than-19500-mother-tongues-spoken-in-india-census-5241056/>



Natural Language Processing

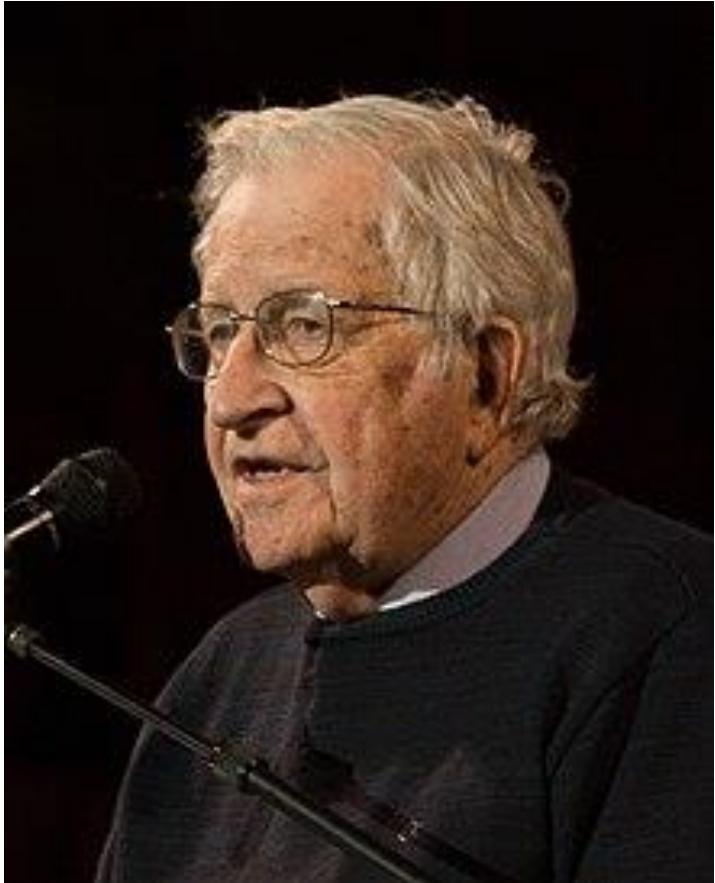


- Setup
 - Two rooms, two humans, and a computer.
 - Room 1: One human C
 - Room 2: One computer (A) and one human (B)
- A response generated from room 2 (either by A or B)
- C has to figure out the source of the response
 - If C is successful → "A" failed the turing test
 - Else, → "A" passed the turing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



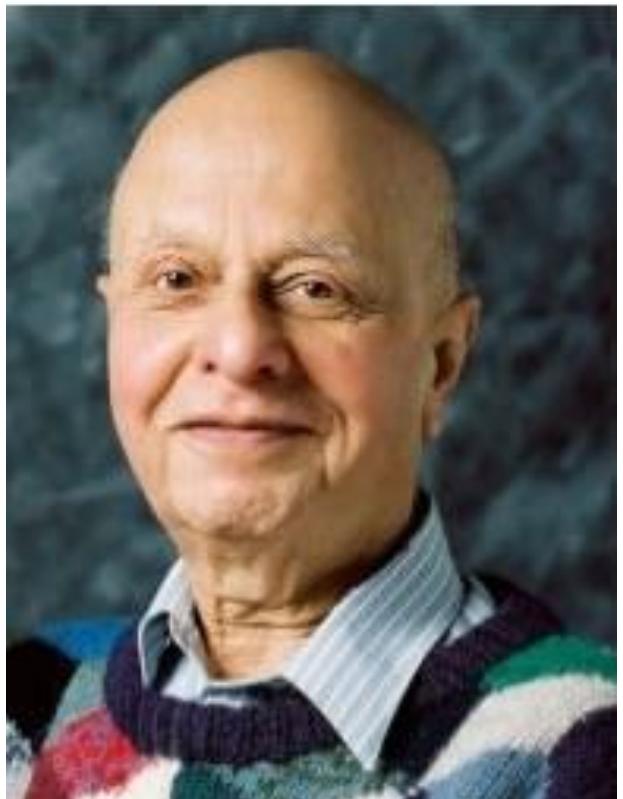
"Computing Machinery and Intelligence" which proposed what is now called the Turing test

Natural Language Processing



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

Natural Language Processing



Aravind Krishna Joshi (August 5, 1929 – December 31, 2017) was a Professor of Computer and Cognitive Science in University of Pennsylvania.

Joshi defined the **tree-adjoining grammar formalism** which is often used in computational linguistics and natural language processing.

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 psychologist, rephrasing her response with a few grammar rules. ^[7]
SHRDLU	1970	Terry Winograd	a natural language system worked on related "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	
Jabberwocky	1982	Rollo Carpenter	chatterbot with stated aim to "simulate natural human chat in an interesting, entertaining and humorous manner".
MUMBLE (software)	1982	McDonald et al.	
Racter	1983	William Chamberlain and Thomas Etter	chatterbot that generated English language prose at random.
MOPTRANS ^[8]	1984	Lehnert	
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.

Why NLP is challenging?

Ambiguity

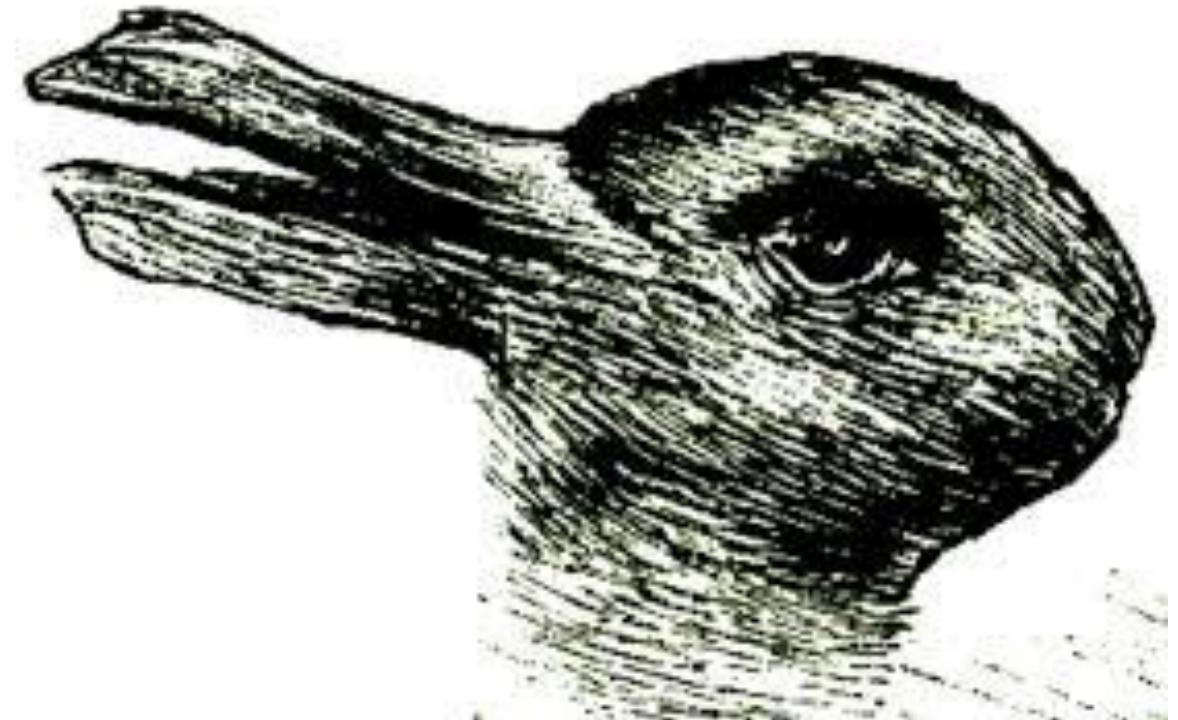
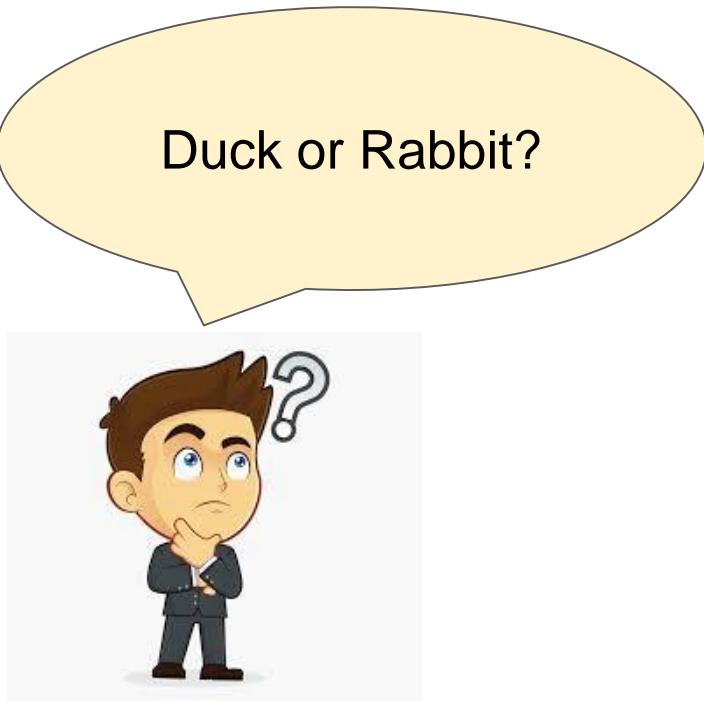
The real reason why NLP is hard

“Rohit Sharma was **on fire** last night. He totally **destroyed** the other teams”



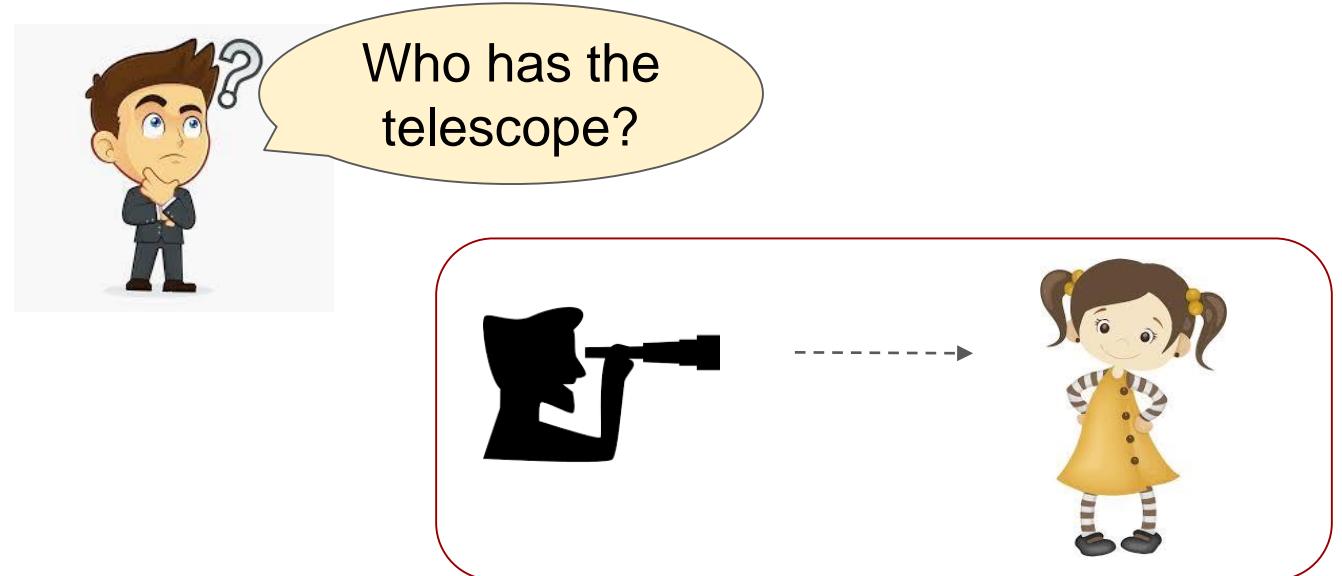
Ambiguity

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

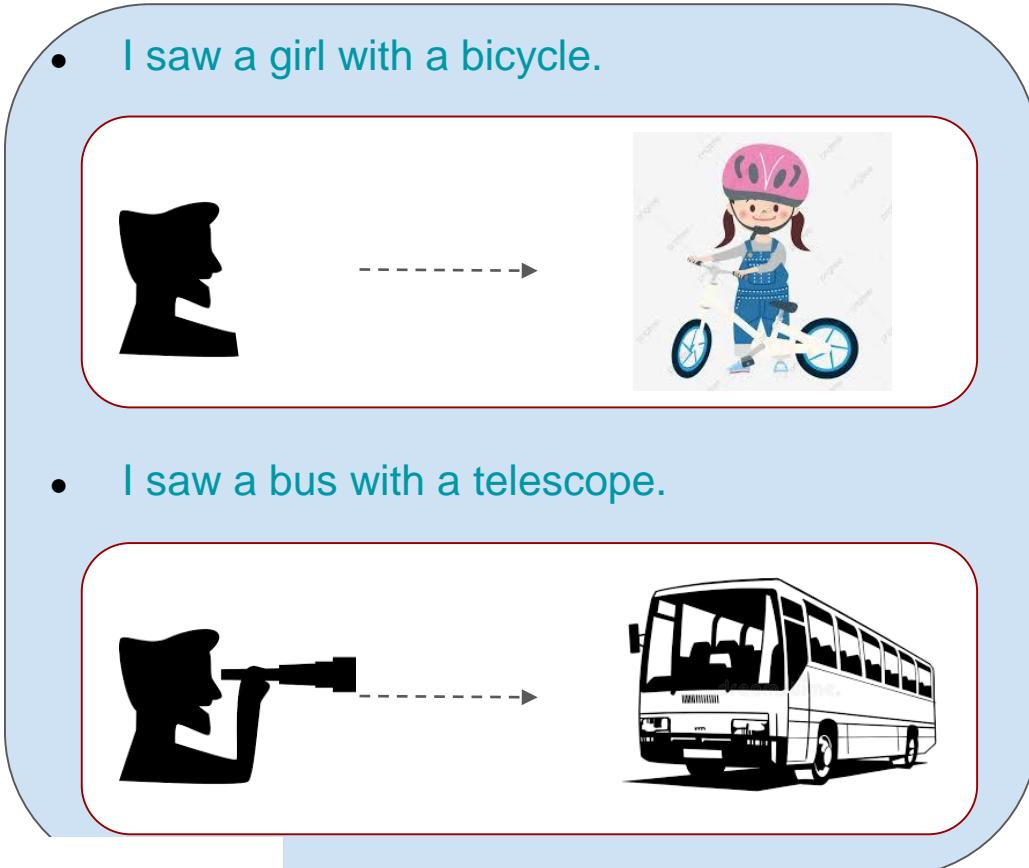


Ambiguity in language

- I saw a girl with a telescope.



- I saw a girl with a bicycle.



- I saw a bus with a telescope.

No ambiguity!



OR

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 padegi.



I have to gift you some sweets.

OR

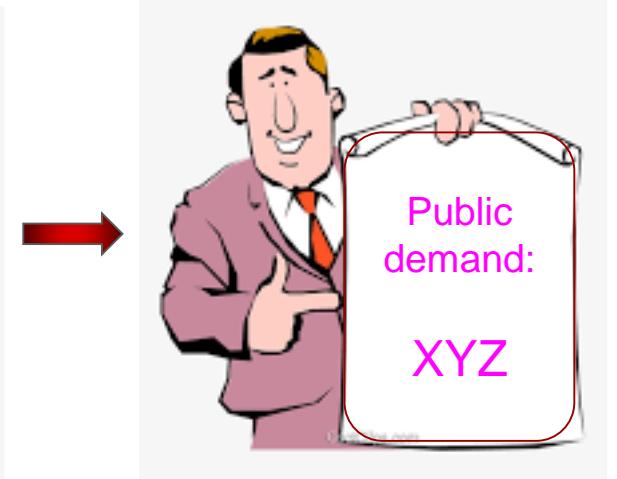
You have to gift me some sweets.

Ambiguity in language

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



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

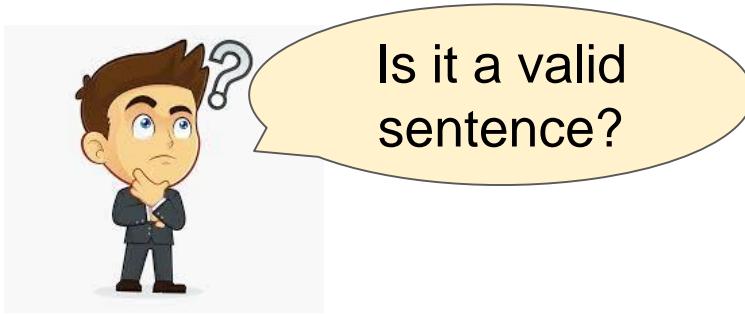
- Violinist Linked to JAL Crash Blossoms => main verb?

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*



The sentence uses a [restrictive clause](#), so there are no commas, nor is there the word "which," as in, "Buffalo *buffalo*, which Buffalo *buffalo* **buffalo**, **buffalo** Buffalo *buffalo*." This clause is also a [reduced relative clause](#), so the word *that*, which could appear between the second and third words of the sentence, is omitted.

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
the New York ... 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 Sue are sisters.
Mary and Sue are mothers.

tricky entity names

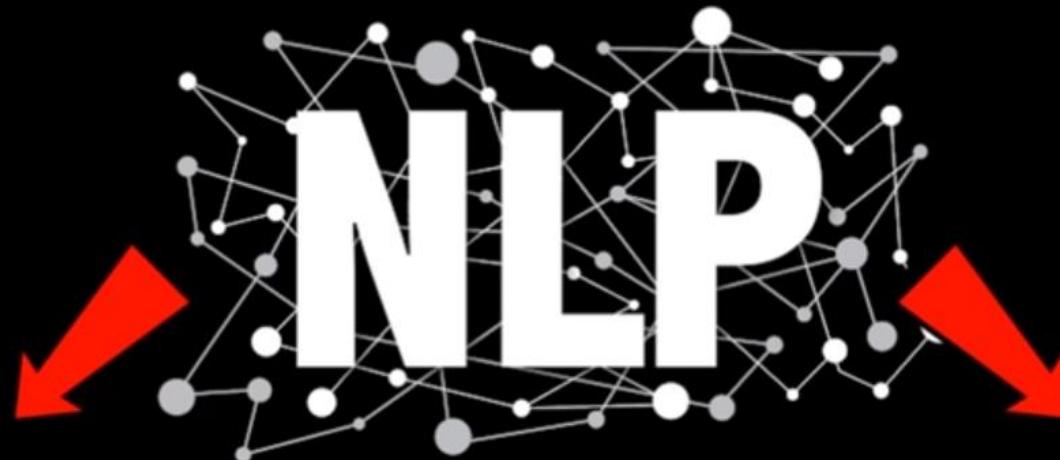
Where is *A Bug's Life* playing ...
Let It Be was recorded ...
... a mutation on the *for* gene ...

that's what makes it fun!

Components of NLP



Natural Language Understanding



Natural Language Generation

NLP layers

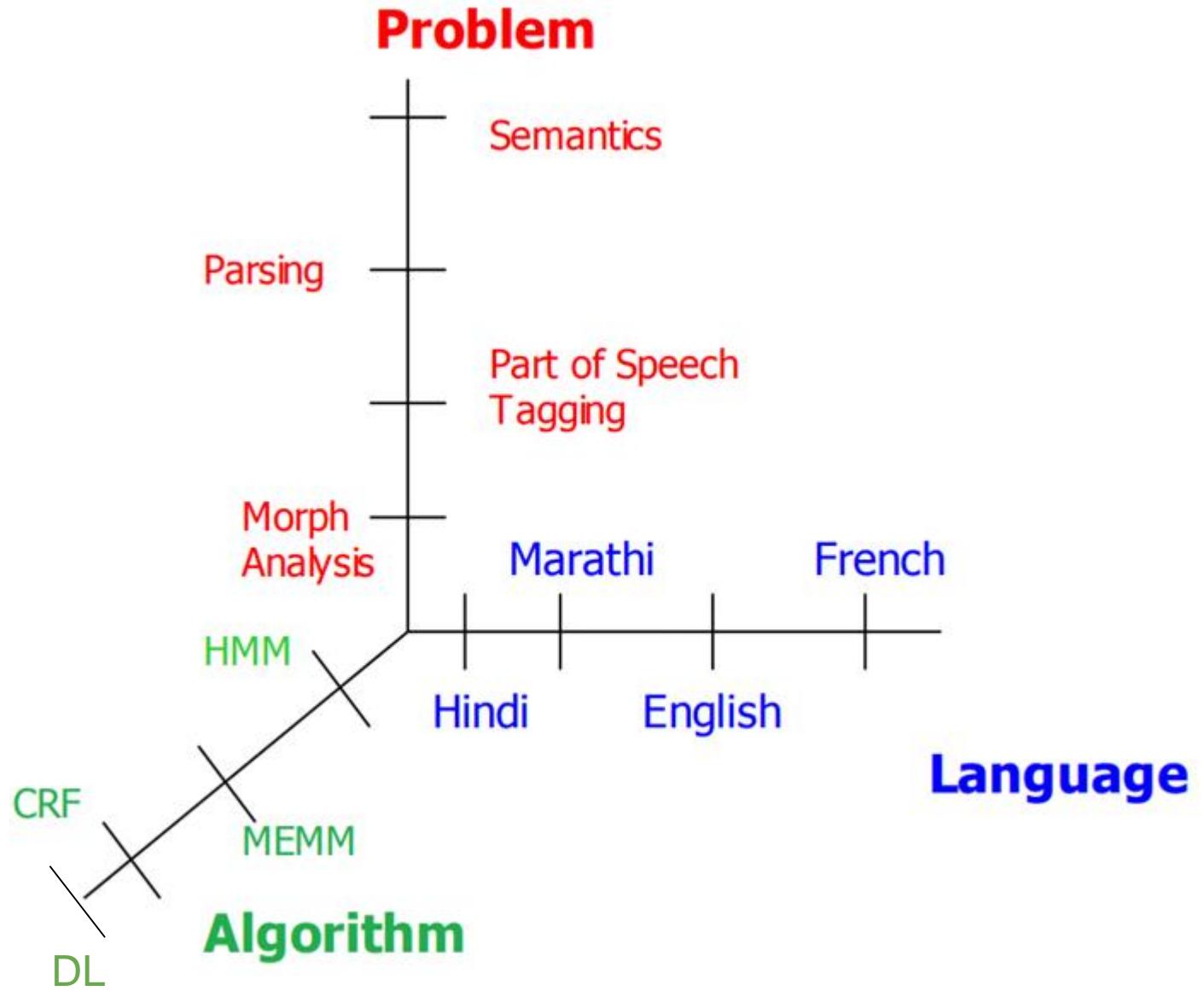
- Understanding the semantics is a non-trivial task.
- Needs to performs a series of incremental tasks to achieve this.
- NLP happens in layers

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](https://www.abc.com)] [:] [/] [/] [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।	
We will go.	हम जाएंगे।	
You will go.	तुम जाओगे।	
He will go.	वह जाएगा।	Different morphological forms of word ‘will go’ in Hindi
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	park	in	a	car	.
PRP	VBD	TO	DT	NN	IN	DT	NN	.

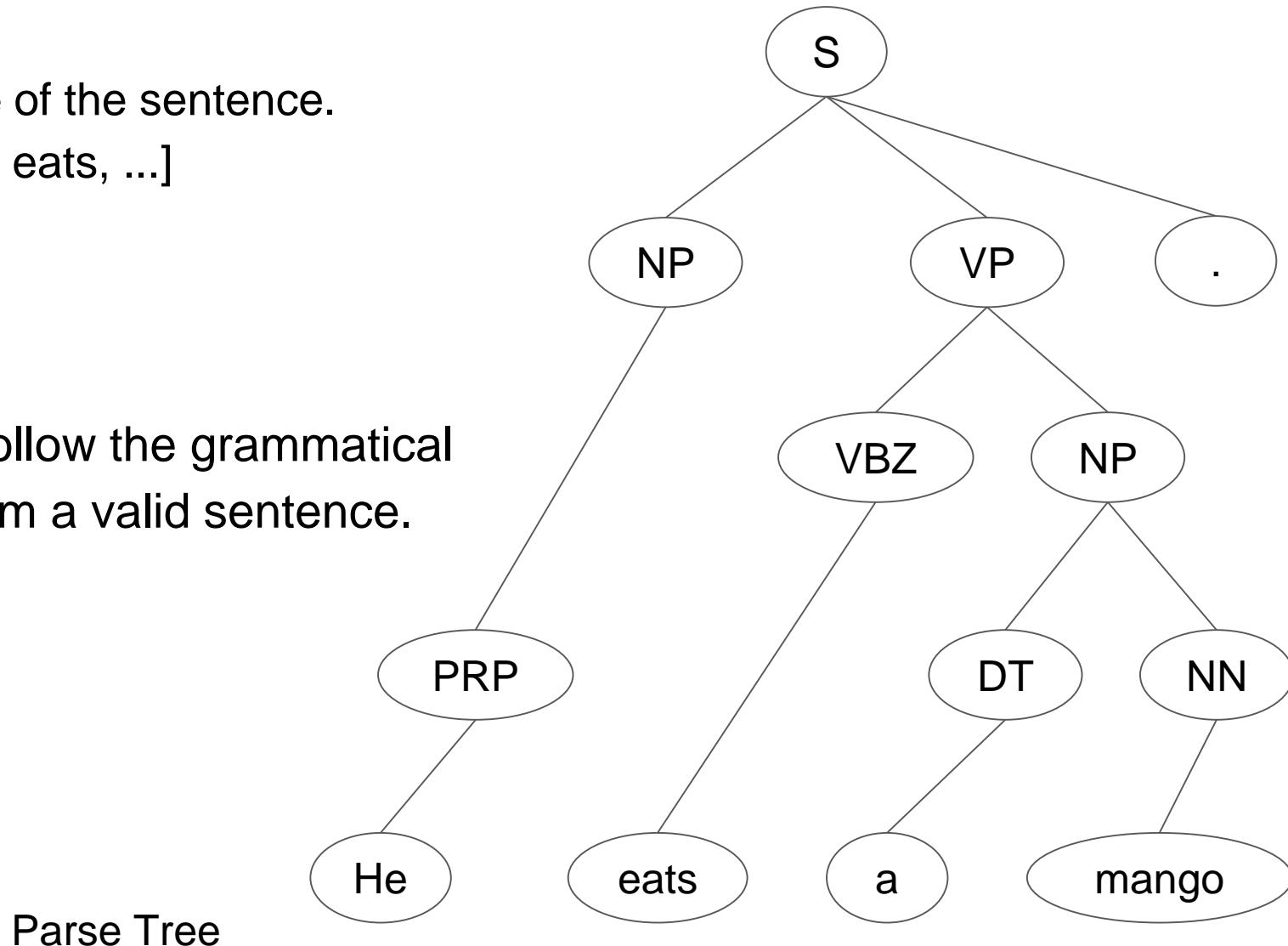
They	went	to	park	the	car	in	the	shed	.
PRP	VBD	TO	VB	DT	NN	IN	DT	NN	.

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.
- Let, vocabulary = [the, mango, he, eats, ...]
 - He eats a mango. $\Rightarrow \checkmark$
 - He mango eats a. $\Rightarrow \times$
- The sequence of words must follow the grammatical structure of the language to form a valid sentence.
 - Construct a parse tree.

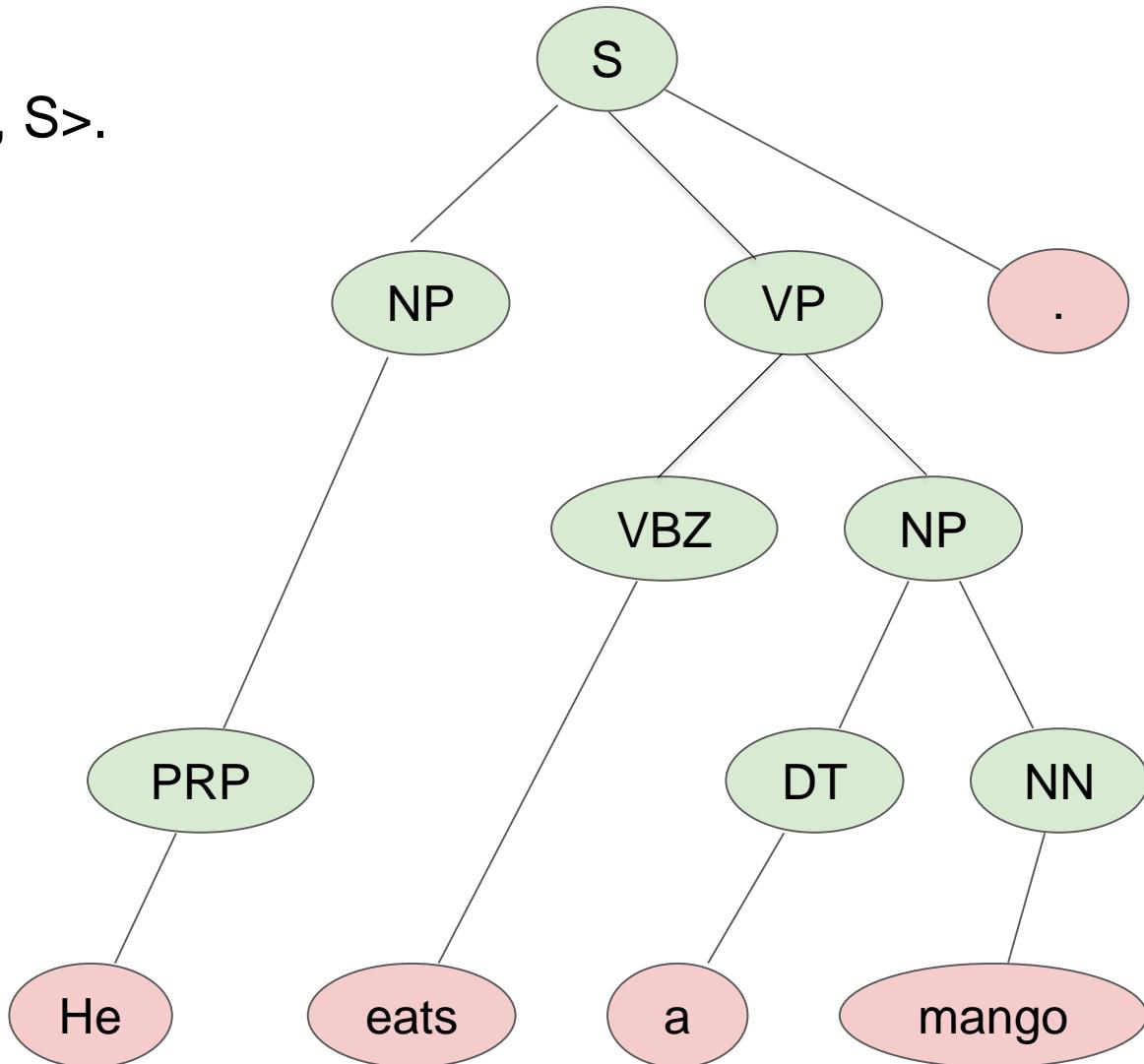


Syntax Processing

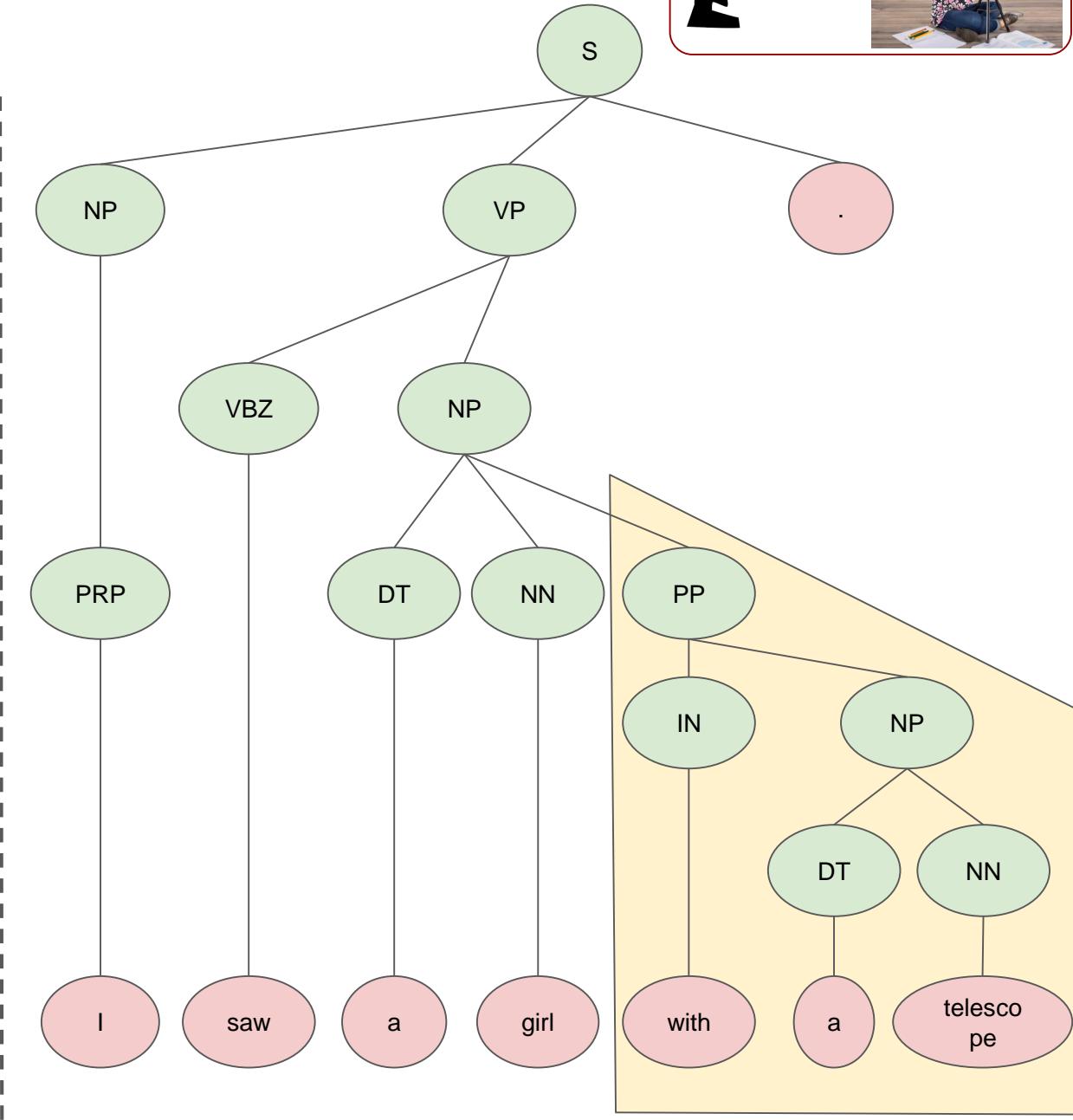
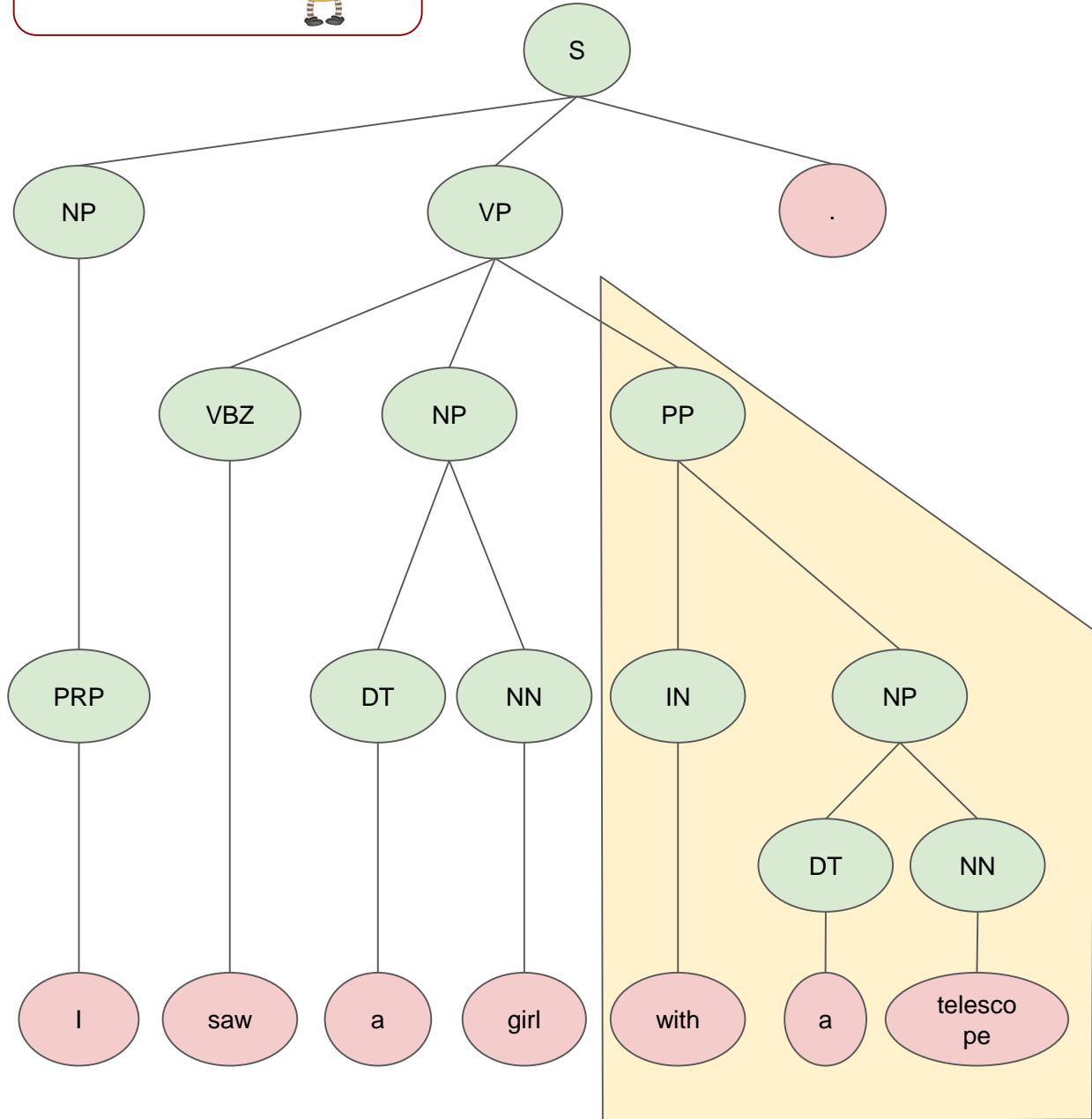
- Every language has a grammar $G = \langle V, T, P, S \rangle$.

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



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 hospitalized*) ⇒ 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.

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

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

The diagram illustrates coreference resolution with two sentences. The first sentence is "John shows Bob his Toyota yesterday. It's similar to the one I bought five years ago." and the second is "That was really nice, but he like this *one* even better." Red curved arrows point from the pronouns "his" and "he" in the first sentence to the noun "one" in the second sentence. A blue curved arrow points from the noun "Toyota" in the first sentence to the noun "one" in the second sentence.

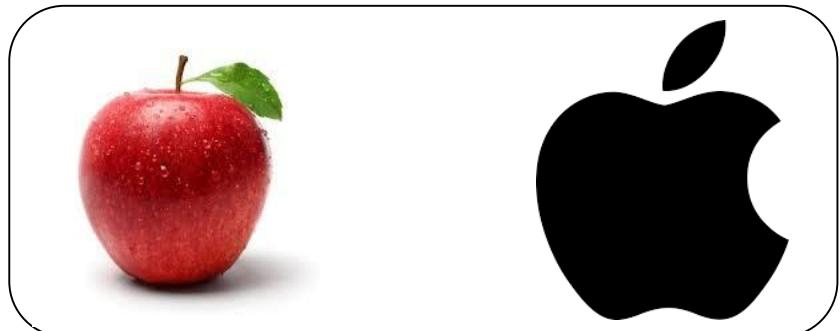
Information Extraction

- 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 ENC 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 doktar 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.

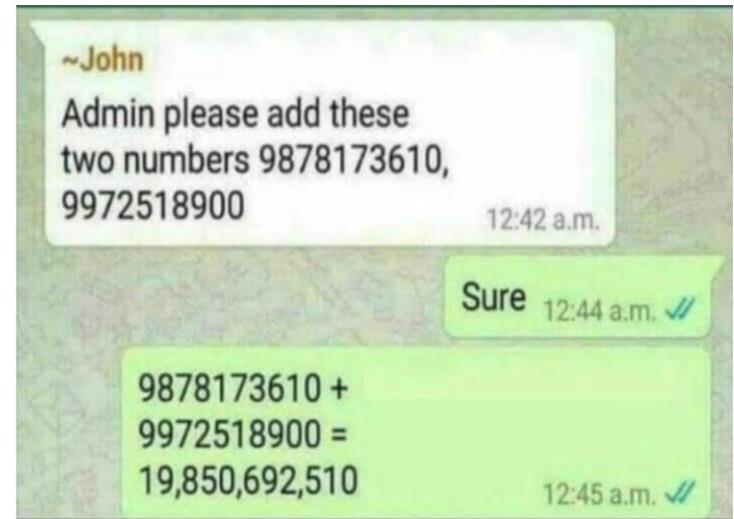
Question Answering

- 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.

Hate Speech

- *Any post that targets a specific individual/group of people based on their ethnicity, religious beliefs, geographical belonging, race, etc., with malicious intentions of disseminating hate or emboldening violence.*
 - *#BuildThatWall #BuildTheDamnWall I'm sorry my Lord #Jesus but people are just deaf down here*
 - *Women ... Can't live with them...Can't shoot them*
- Related terms
 - Insult, Abuse, Offensive, Provocative

Fake News

- A piece of information or an alleged claim that is verifiable to be false.
- Intentionally created posts to spread malicious and false narratives
 - Leverages the chaos/misinformation to gain political, financial, or regional advantages in a quick time

WHO claimed that asymptomatic patients do not spread Covid-19?



Bharat Biotech VP taking the first dose of 'COVAXIN' developed by his team?



INDUSTRY

Ramdev's Patanjali Ayurved claims it can "cure COVID-19 in 7 days"

Cannot cure COVID-19 in 7 days?

Patanjali did not claim?

Language Technology

Mostly solved

Spam detection

Let's go to Agra!



Buy V1AGRA ...



Part-of-speech (POS) tagging

ADJ ADJ NOUN VERB ADV

Colorless green ideas sleep furiously.

Named entity recognition (NER)

PERSON ORG LOC

Einstein met with UN officials in Princeton

Making good progress

Sentiment analysis

Best roast chicken in San Francisco!



The waiter ignored us for 20 minutes.



Coreference resolution

Carter told Mubarak he shouldn't run again.

Word sense disambiguation (WSD)

I need new batteries for my **mouse**.



Parsing



I can see Alcatraz from the window!

Machine translation (MT)

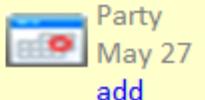
第13届上海国际电影节开幕...



The 13th Shanghai International Film Festival...

Information extraction (IE)

You're invited to our dinner party, Friday May 27 at 8:30



Still really hard

Question answering (QA)

Q. How effective is ibuprofen in reducing fever in patients with acute febrile illness?

Paraphrase

XYZ acquired ABC yesterday

ABC has been taken over by XYZ

Summarization

The Dow Jones is up

The S&P500 jumped

Housing prices rose



Economy is good

Dialog

Where is Citizen Kane playing in SF?



Castro Theatre at 7:30. Do you want a ticket?



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.