

Introduction to NLP

Bal Krishna Nyaupane

Assistant Professor

Department of Electronics and Computer Engineering

Paschimanchal Campus, IOE

bkn@wrc.edu.np

Dreams ??



Introduction to NLP

- NLP is the branch of computer science focused on developing systems that allow computers to communicate with people using everyday language.
- NLP is the use of computers to process written and spoken language for some practical useful purpose.
- NLP refers to AI method of communicating with an intelligent systems using a natural language such as English.
- Processing of Natural Language is required when you want an intelligent system like robot to perform as per your instructions, when you want to hear decision from a dialogue based clinical expert system, etc.
- The field of NLP involves making computers to perform useful tasks with the natural languages humans use. The input and output of an NLP system can be – Speech and Written Text.

What is NLP

- An experimental computer science research area that includes problems and solutions pertaining to the understanding of human language.
- *Wiki:* Natural language processing(NLP) is a field of computer science, artificial intelligence, and computational linguistics concerned with the interactions between computers and human (natural) languages.

NLP Terminology: Morpheme

- A "morpheme" is a short segment of language that meets three basic criteria:
 1. *It is a word or a part of a word that has meaning.*
 2. *It cannot be divided into smaller meaningful segments without changing its meaning or leaving a meaningless remainder.*
 3. *It has relatively the same stable meaning in different verbal environments.*
- ***Free and Bound Morphemes***
 - "Free morphemes" can stand alone with a specific meaning, for example, eat, date, weak.
 - "Bound morphemes" cannot stand alone with meaning. Morphemes are comprised of two separate classes called (a) bases (or roots) and (b) affixes.

NLP Terminology: Morpheme

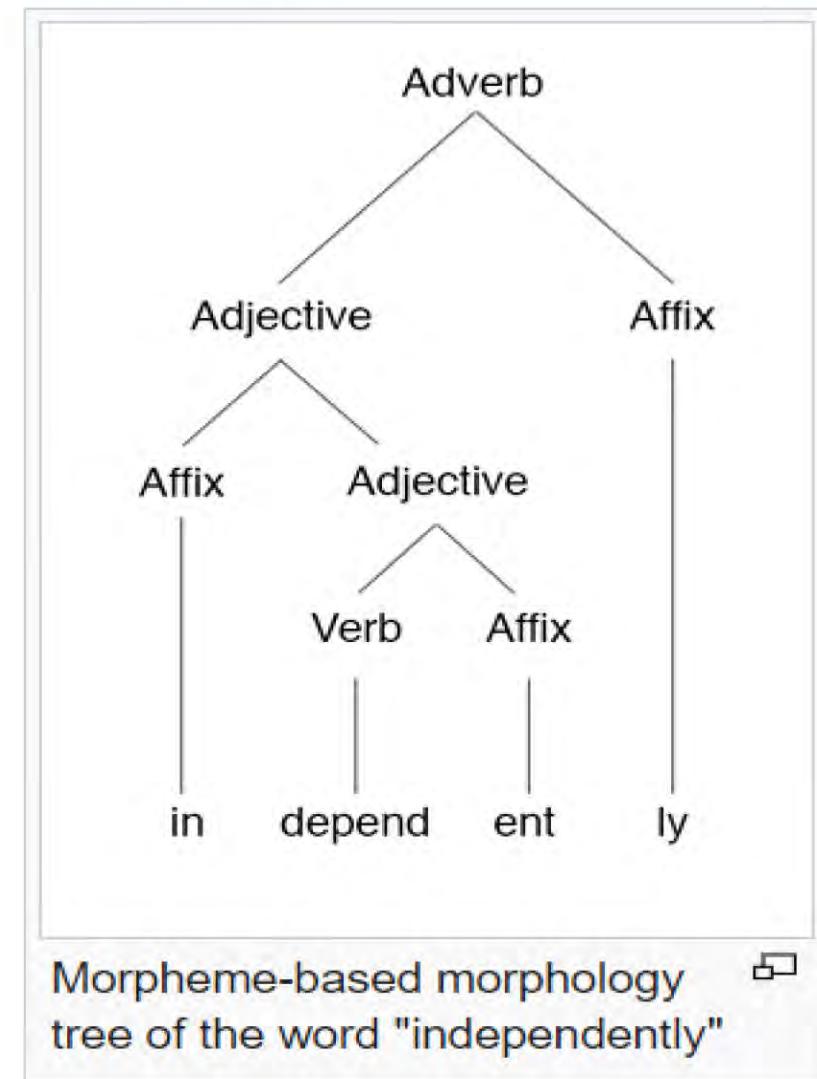
- A "base," or "root" is a morpheme in a word that gives the word its principle meaning. An example of a "***free base***" ***morpheme is woman*** in the ***word womanly***.
- **Affixes:** An "affix" is a bound morpheme that occurs before or after a base.
 - An affix that **comes before a base** is called a "**prefix**". Examples: ante-, pre-, un-, and dis-, as in the following words: antedate, prehistoric, unhealthy, disregard etc.
 - An affix that **comes after a base** is called a "**suffix**". Examples: -ly, -er, -ism, and -ness, as in the following words: happily, gardener, capitalism, kindness.
- **Derivational Affixes:** "Derivational affixes" serve to alter the meaning of a word by building on a base. The addition of the **prefix un-** to **healthy** alters the meaning of **healthy**. The **suffix -er** to **garden** changes the meaning of **garden**.

NLP Terminology: Morpheme

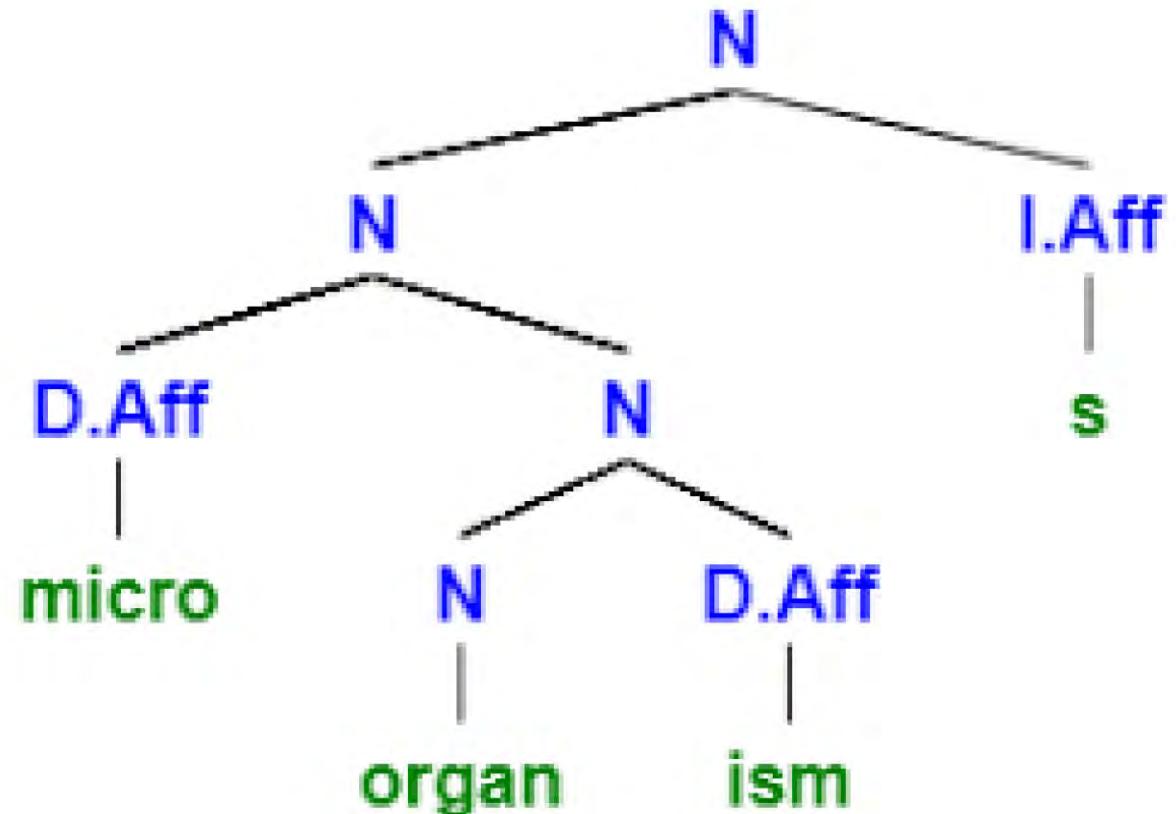
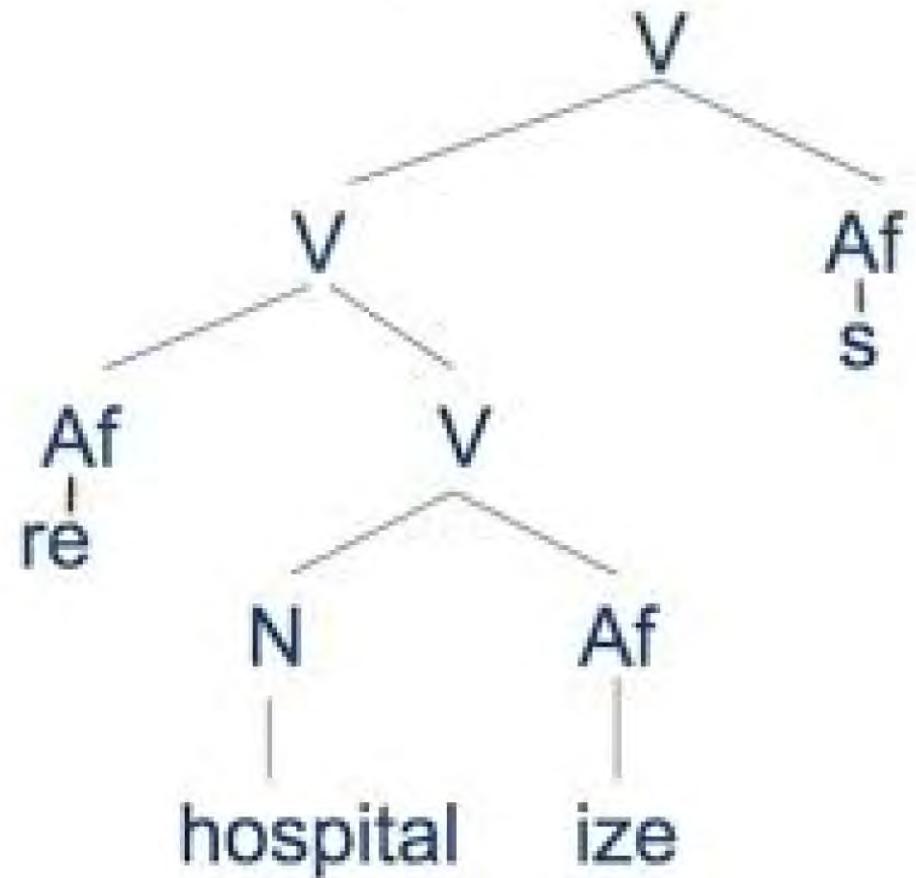
- **Inflectional Affixes:** there are only eight "inflectional affixes" in English, and these are all suffixes. These grammatical functions are shown to the right of each suffix.
 - *-s noun plural*
 - *'s noun possessive*
 - *-s verb present tense third person singular*
 - *-ing verb present participle/gerund*
 - *-ed verb simple past tense*
 - *-en verb past perfect participle*
 - *-er adjective comparative*
 - *-est adjective superlative*

NLP Terminology: Morphology

- Morphology is the study of word structure, the way words are formed and the way their form interacts with other aspects of grammar such as phonology and syntax.
- Example:
 - carried* ⇒ *carry* + *ed* (past tense)
 - independently* ⇒ *in* + (*depend* + *ent*) + *ly*
 - Googlers* ⇒ (*Google* + *er*) + *s* (plural)



NLP Terminology: Morphology



NLP Terminology: Syntax

- Syntax defines rules, principles, processes that govern sentence structure of a language. Syntax can differ widely among languages, but every language has systematic structural principles.
- Syntax determines the ordering of Subject, Verb, and Object in a sentence.

Word order	English equivalent	Proportion of languages	Example languages
SOV	"She him loves."	45%	Hindi, Latin, Japanese, Marathi
SVO	"She loves him."	42%	English, Hausa, Mandarin, Russian
VSO	"Loves she him."	9%	Biblical Hebrew, Irish, Filipino, Tuareg
VOS	"Loves him she."	3%	Malagasy, Baure
OVS	"Him loves she."	1%	Apalaí, Hixkaryana
OSV	"Him she loves."	0%	Warao

Frequency distribution of word order in languages surveyed by Russell S. Tomlin in 1980s^{[1][2]} (v · T · E)

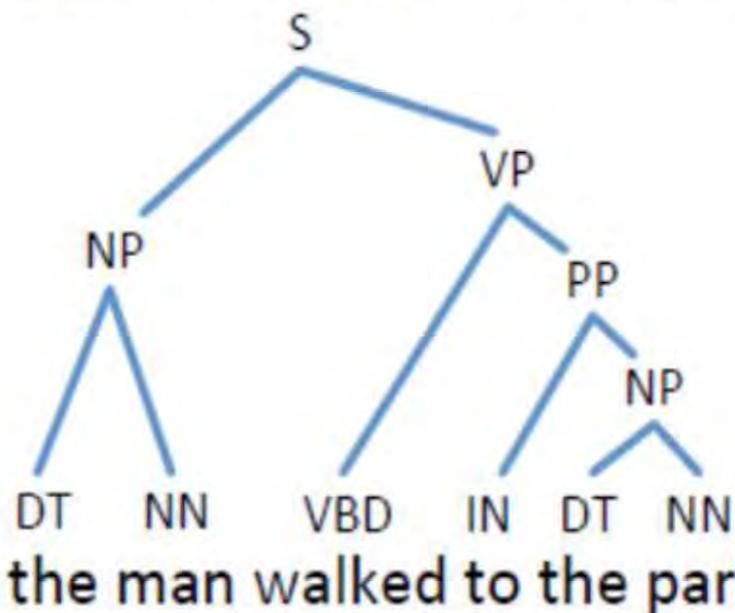
NLP Terminology: Syntax

- **Grammars:** We will use grammar to denote a formal object that represents the rules/principles/processes that determine sentence structure.
 - focuses on the **constituent** relation
 - informally: “sentences have hierarchical structure”
 - a sentence is made up of two pieces:
 - subject, typically a **noun phrase (NP)**
 - predicate, typically a **verb phrase (VP)**
 - NPs and VPs are in turn made of up of pieces:
 - old books = (old + books)
 - the old books = (the + (old + books))
 - walked to the park = (walked + (to + (the + park))))
 - each parenthesized phrase is a **constituent** in the **constituent parse**

NLP Basic Terminology: Syntax

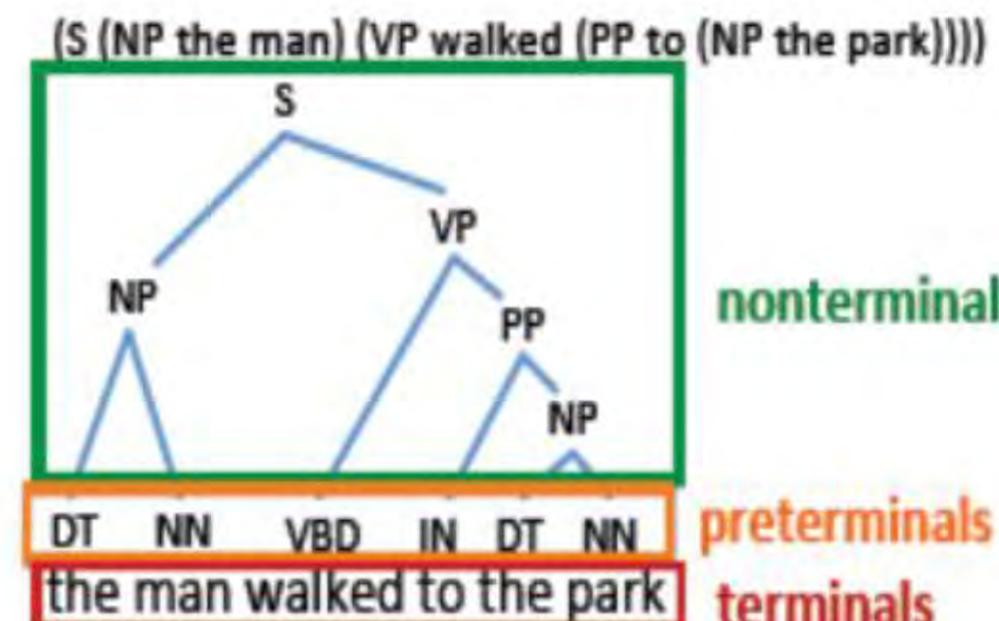
- Constituent parse = bracketing (that represents the hierarchical structure)
 - e.g., sentence: the man walked to the park
 - bracketing : *((the man) (walked (to (the park))))*
- Bracketing → Tree

(S (NP the man) (VP walked (PP to (NP the park))))



Key:

S = sentence
NP = noun phrase
VP = verb phrase
PP = prepositional phrase
DT = determiner
NN = noun
VBD = verb (past tense)
IN = preposition



NLP Terminology: Syntactic Ambiguities

- *Prepositional phrase attachment ambiguity*



NLP Terminology: Syntactic Ambiguities

- *Coordination ambiguities*: often found when modifiers are used with conjunctions.
 - keyboard and monitor with the Apple logo
 - (keyboard and monitor) with the Apple logo
 - keyboard and (monitor with the Apple logo)
 - old men and women
 - old (men and women)
 - (old men) and women

NLP Terminology: Semantics

- *Semantics is concerned with the meaning of words and how to combine words into meaningful phrases and sentences.*
- Semantic analysis describes the process of understanding natural language—the way that humans communicate—based on meaning and context.
- Semantic analysis is the process of finding the meaning from text. This analysis gives the power to computers to understand and interpret sentences, paragraphs, or whole documents, by analyzing their grammatical structure, and identifying the relationships between individual words of the sentence in a particular context.
- Therefore, the goal of semantic analysis is *to draw exact meaning or dictionary meaning from the text.* The work of a semantic analyzer is to check the text for meaningfulness.

NLP Terminology: Semantic Analysis

- **How is Semantic Analysis different from Lexical Analysis?**
 - Lexical analysis is based on *smaller tokens* but on the contrary, the semantic analysis *focuses on larger chunks*.
- **Studying meaning of individual word:** It is the first part of the semantic analysis in which the study of the meaning of individual words is performed. This part is called lexical semantics.
- **Studying the combination of individual words:** In the second part, the individual words will be combined to provide meaning in sentences.
- For example, analyze the sentence “**Ram is great.**” In this sentence, the speaker is talking either about Lord Ram or about a person whose name is Ram.

NLP Terminology: Semantic Analysis

■ Elements of Semantic Analysis

- ***Hyponymy:*** It represents the relationship between a generic term and instances of that generic term e.g. orange is a hyponym of fruit (hypernym).
- ***Homonyms:*** It may be described as words with the same spelling or form but diverse and unconnected meanings e.g. orange (color), orange (fruit).
- ***Polysemy:*** Polysemy has the same spelling but various and related meanings.
- **Difference between Polysemy and Homonymy**
 - The “Bank” word is a polysemy word having the following meanings:
 - ✓ *A financial institution.*
 - ✓ *The building in which such an institution is located.*
 - ✓ *A synonym for “to rely on”.*

NLP Basic Terminology: Semantic Analysis

■ Elements of Semantic Analysis

- For Example, the same word “Bank”, we can write the meaning as
 - ✓ ‘*a financial institution*’ or
 - ✓ ‘*a river bank*’.
- In that case, the meanings are unrelated to each other.
- **Synonymy:** It is the relation between two lexical items having different forms but expressing the same or a close meaning. e.g., ‘author/writer’, ‘fate/destiny’.
- **Antonyms:** words that have close to opposite meanings e.g., 'happy/sad', 'rich/poor', 'hot/cold', 'life/death'.
- **Meronymy:** a logical arrangement of text and words that denotes a constituent part of or member of something e.g., a segment of an orange.

NLP Terminology: Processes of Semantic Analysis

- **Word Sense Disambiguation:** In semantic analysis with machine learning, computers use Word Sense Disambiguation to determine which meaning is correct in the given context.
- **Example:** Consider the word: “Orange”
 - Can refer to a color, a fruit, or even a city in Florida!
- **Example:** Mary had a bat in her office.
 - bat = ‘a baseball thing’ or bat = ‘a flying mammal’



NLP Terminology: Processes of Semantic Analysis

- **Relationship Extraction:** try to detect the semantic relationships present in a text. Usually, relationships involve two or more entities such as names of people, places, company names, etc.
- For Example: Consider the following phrase
"Steve Jobs is the founder of Apple, which is headquartered in California."
- The above phrase contains two different relationships:

Steve Jobs founder of Apple.

[Person]

[Company]

Apple headquartered in California.

[Company]

[Place]

NLP Terminology: Semantic Analysis Techniques

- ***Topic classification:*** sorting text into predefined categories based on its content.
- ***Sentiment analysis:*** It is a method for detecting the hidden sentiment inside a text, may it be positive, negative or neutral. This method helps in understanding the urgency of any statement. In social media, often customers reveal their opinion about any concerned company.
- ***Intent classification:*** It is a method of differentiating any text on the basis of the intent of your customers. The customers might be interested or disinterested in your company or services. Knowing prior whether someone is interested or not helps in proactively reaching out to your real customer base.

NLP Terminology: Semantic Analysis Techniques

- ***Keyword Extraction:*** It is a method of extracting the relevant words and expressions in any text to find out the granular insights. It is mostly used along with the different classification models. It is used to analyze different keywords in a corpus of text and detect which words are ‘negative’ and which words are ‘positive’. The topics or words mentioned the most could give insights of the intent of the text.
- ***Entity extraction:*** identifying named entities in text, like names of people, companies, places, etc. A customer service team might find this useful to automatically extract names of products, shipping numbers, emails, and any other relevant data from customer support tickets.

■ **Discourse integration**

- This analysis deals with how the immediately preceding sentence can affect the meaning and interpretation of the next sentence.
- Here, context can be analyzed in a bigger context, such as paragraph level, document level, and so on.
- For example, interpreting pronouns and interpreting the temporal aspects of the information.
- For example, the word “that” in the sentence “He wanted that” depends upon the prior discourse context.

NLP Terminology

- **Pragmatics:** It deals with using and understanding sentences in different situations and how the interpretation of the sentence is affected.
- **World Knowledge:** It includes the general knowledge about the world. What each language user must know about the other's beliefs and goals.

Syntax, Semantic, Pragmatics

- Syntax concerns the proper ordering of words and its affect on meaning.
 - The dog bit the boy.
 - The boy bit the dog.
 - Bit boy dog the the.
 - Colorless green ideas sleep furiously.
- Semantics concerns the (literal) meaning of words, phrases, and sentences.
 - “plant” as a photosynthetic organism
 - “plant” as a manufacturing facility
 - “plant” as the act of sowing
- Pragmatics concerns the overall communicative and social context and its effect on interpretation.
 - The ham sandwich wants another beer. (co-reference, anaphora)

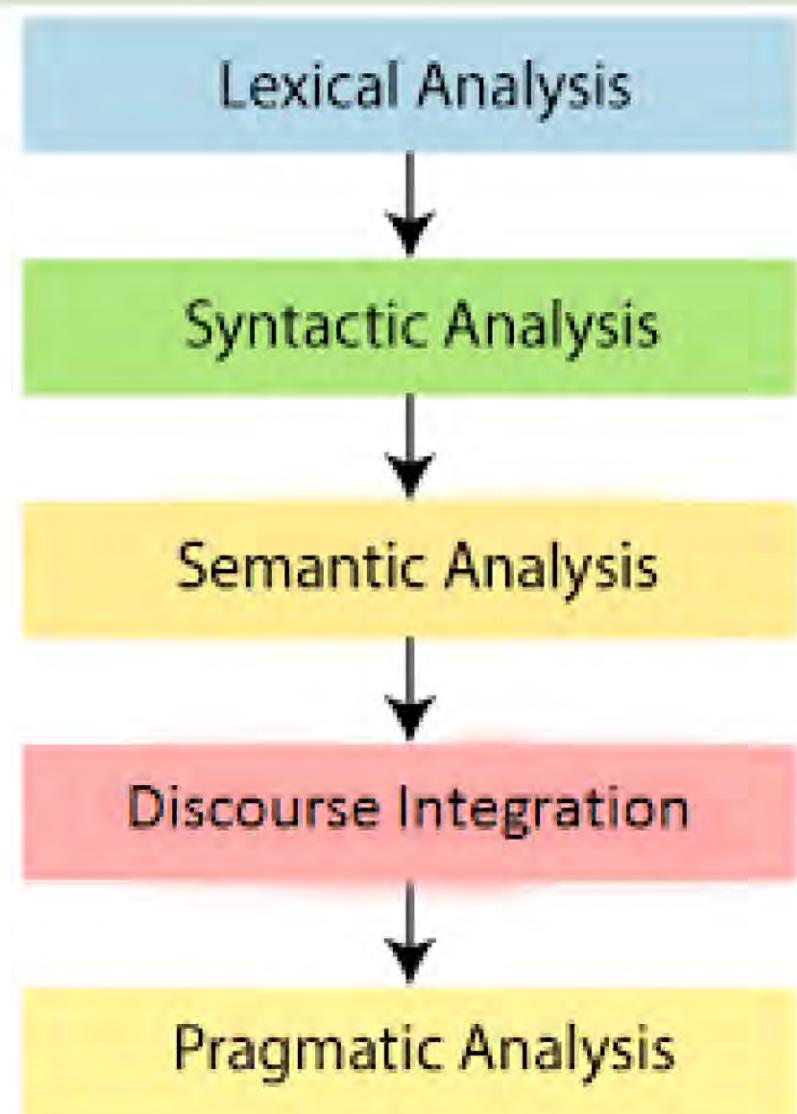
Steps in NLP

■ Lexical Analysis

- It involves identifying and analyzing the structure of words. Lexicon of a language means the collection of words and phrases in a language.
- Lexical analysis is dividing the whole chunk of text into paragraphs, sentences, and words.

■ Syntactic Analysis (Parsing)

- It involves analysis of words in the sentence for grammar and arranging words in a manner that shows the relationship among the words.
- It takes an input sentence and produces a representation of its grammatical structure.



■ Implementation Aspects of Syntactic Analysis

- **Context-Free Grammar:** It is the grammar that consists rules with a single symbol on the left-hand side of the rewrite rules.
- A **parse tree** breaks down the sentence into structured parts so that the computer can easily understand and process it. Each level of the tree corresponds to the application of one grammar rule. In order for the parsing algorithm to construct this parse tree, a set of rewrite rules, which describe what tree structures are legal, need to be constructed.
- **Grammar rule:**

S → NP VP

Articles (DET) → a | an | the,

Nouns (N) → bird | birds | grain | grains

Steps in NLP

- Grammar rule:

Noun Phrase (NP) → Article + Noun | Article + Adjective + Noun

Verbs (V) → pecks | pecking | pecked

Verb Phrase (VP) → NP V | V NP

Adjectives (ADJ) → beautiful | small | chirping

sentence → **noun_phrase, verb_phrase**

noun_phrase → **proper_noun**

noun_phrase → **determiner, noun**

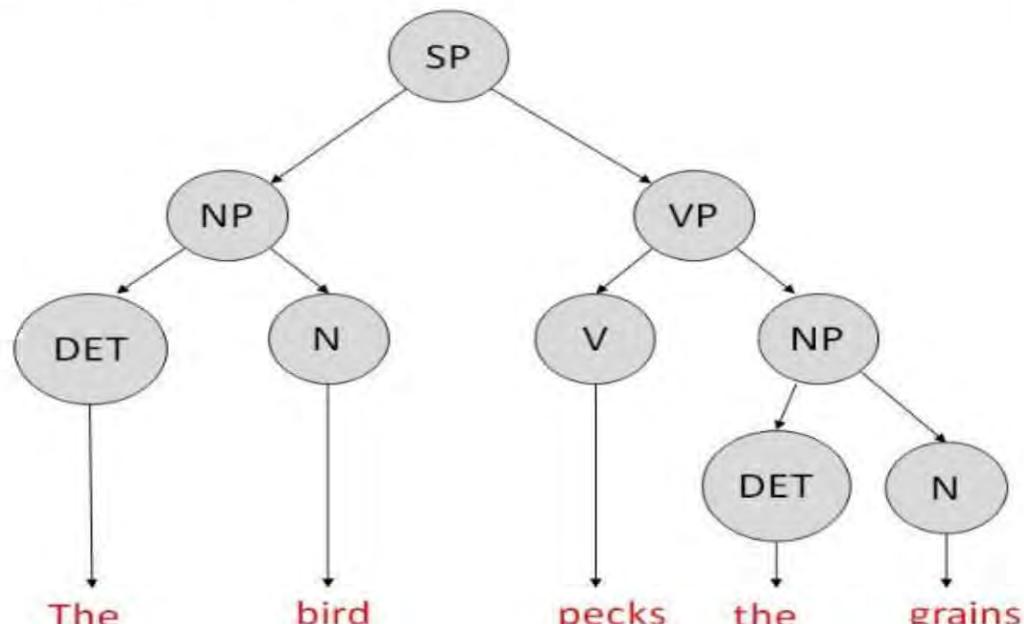
verb_phrase → **verb, noun_phrase**

proper_noun → [mary]

noun → [apple]

verb → [ate]

determiner → [the]



- **Semantic Analysis**
 - It draws the exact meaning or the dictionary meaning from the text.
 - It is done by mapping syntactic structures and objects in the task domain.
- **Discourse Integration**
 - The meaning of any sentence depends upon the meaning of the sentence just before it. In addition, it also brings about the meaning of immediately succeeding sentence.
- **Pragmatic Analysis**
 - During this, what was said is re-interpreted on what it actually meant. It involves deriving those aspects of language which require real world knowledge.

Steps in NLP

- Pragmatics comprises aspects of meaning that depend upon the context or upon facts about real world. These aspects include:
 - *Pronouns and referring expressions*
 - *Logical inferences* that can be drawn from the meanings of a set of propositions.
 - *Discourse structure*: the meaning of a collection of sentences taken together.
- Example
 - **Handling Pronouns:** “Mary eats apples. She likes them.”
She=“Mary”, them=“apples”.
 - **Handling ambiguity:** Pragmatic ambiguity: “you’re late”.
What’s the speaker’s intention: informing or criticizing?

Steps in NLP

Determine which phrases in a document refer to the same underlying entity.

- John put the **carrot** on the **plate** and ate **it**.
- **Bush** started the war in Iraq. But **the president** needed the consent of **Congress**.

Some cases require difficult reasoning.

- Today was Jack's birthday. Penny and Janet went to the store. They were going to get presents. Janet decided to get a kite. "Don't do that," said Penny. "Jack has a kite. He will make you take it back."

Ellipsis Resolution

- Frequently words and phrases are omitted from sentences when they can be inferred from context.
- "Wise men talk because they have something to say; fools, because they have to say something." (Plato)
 - **talk because they have to say something.**" (Plato)

Applications of NLP: Sentiment Analysis

- Sentiment analysis is able to recognize subtle nuances in emotions and opinions - and determine how positive or negative they are. Sentiment analysis is used to identify the sentiments among several posts.
- It is also used to identify the sentiment where the emotions are not expressed explicitly.
- Companies are using sentiment analysis, an application of natural language processing (NLP) to identify the opinion and sentiment of their customers online.
- It will help companies to understand what their customers think about the products and services. Companies can judge their overall reputation from customer posts with the help of sentiment analysis.

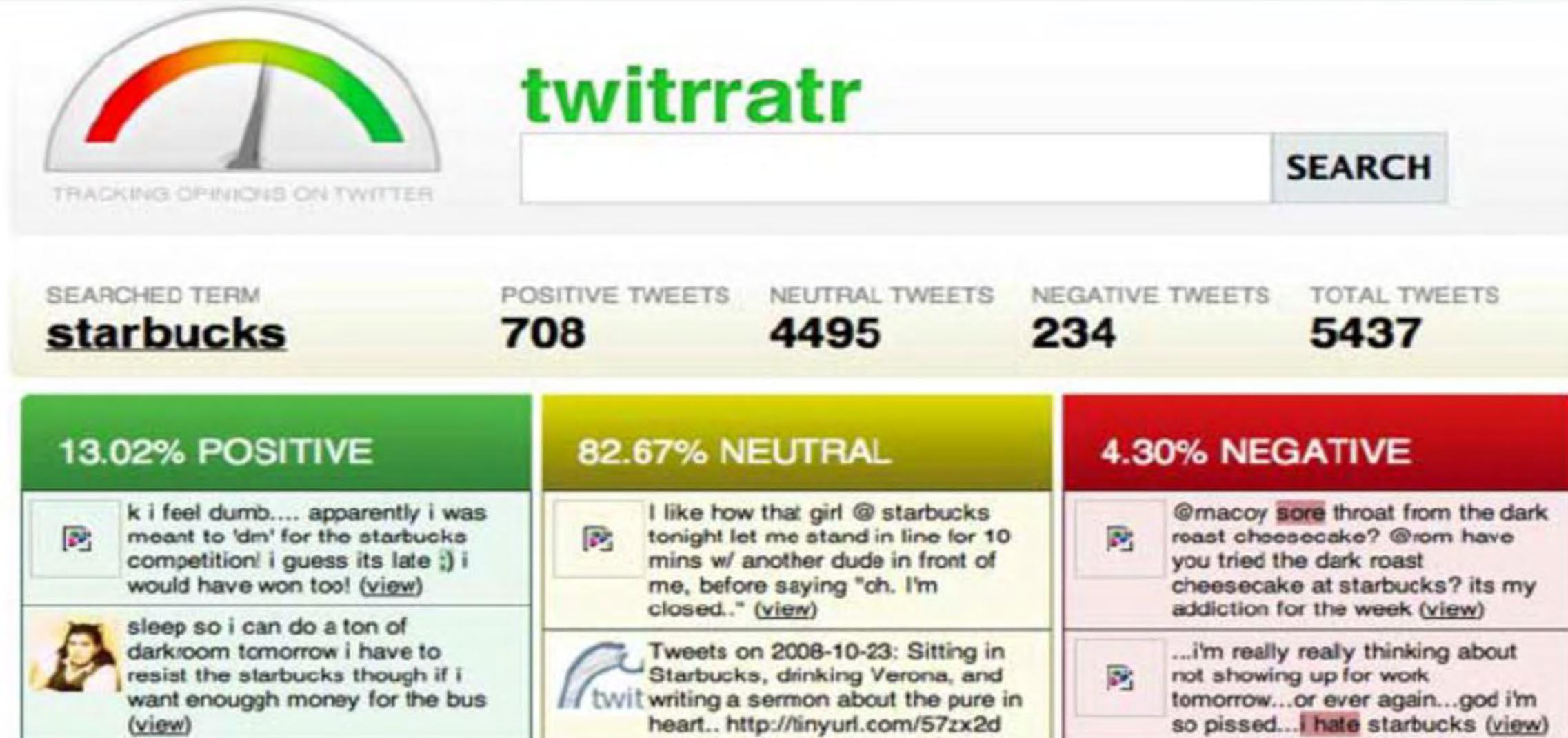
Test with your own text

Their customer service was
terrible. I was on hold for 2
hours!

Results

TAG	CONFIDENCE
Negative	99.9%

Applications of NLP: Sentiment Analysis



Applications of NLP: Sentiment Analysis

- customer review sentiment classification:

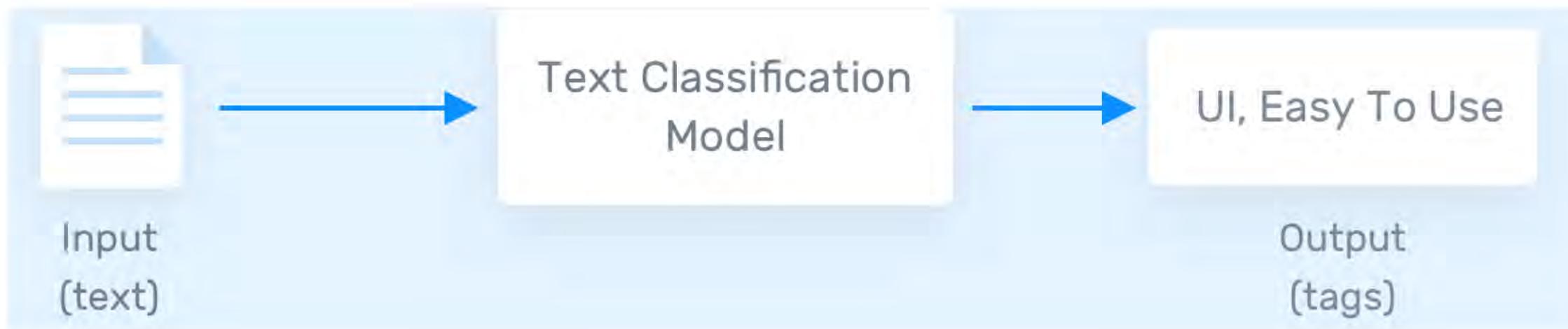
it works with a minimum of fuss .	positive
size - bigger than the ipod	negative
i 've had this thing just over a month and the headphone jack has already come loose .	negative
you can manage your profile , change the contrast of backlight , make different type of display , either list or tabbed .	positive
i replaced it with a router raizer and it works much better .	negative

Applications of NLP: Text Classification

- Text classification allows NLP and AI technologies to automatically understand, process, and categorize unstructured text.
- Text classification is one of the fundamental tasks in natural language processing with broad applications such as sentiment analysis, topic labeling, spam detection, and intent detection.
- Using natural language processing models, data could be sorted and organized based on predetermined tags and categories.

Applications of NLP: Text Classification

- Here's an example of how it works: "The user interface is quite straightforward and easy to use."
 - A text classifier can take this phrase as an input, analyze its content, and then automatically assign relevant tags, such as UI and Easy To Use.



Applications of NLP: Text Classification

The screenshot shows a Gmail inbox with the following categories and their contents:

- Primary**:
 - Google+ (3 new): You were tagged in 3 photos on Google+ - Google+ You were tagged in three pi...
 - YouTube (1 new): LauraBlack just uploaded a video. - Jess, have you seen the video LauraBlack u...
 - Emily Million (Google+) (1 new): [Knitting Club] Are we knitting tonight? - [Knitting Club] Are we knitting tonight?
 - Sean Smith (Google+) (1 new): Photos of the new pup - Sean Smith shared an album with you. View album be tho...
 - Google+ (1 new): Kate Baynham shared a post with you - Follow and share with Kate by adding her -
 - Google+ (1 new): Danielle Hoodhood added you on Google+ - Follow and share with Danielle by -
 - YouTube (1 new): Just for You From YouTube: Daily Update - Jun 19, 2013 - Check out the latest
 - Google+ (1 new): You were tagged in 3 photos on Google+ - Google+ You were tagged in three pho...
 - Hilary Jacobs (Google+) (1 new): Check out photos of my new apt - Hilary Jacobs shared an album with you. View
 - Google+ (1 new): Kate Baynham added you on Google+ - Follow and share with Kate by adding her -
- Social** (3 new): Google+, YouTube, Emi...
- Promotions** (2 new): Google Offers, Zagat
- Updates** (2 new): Shoehop, Blitz Air

On the left, there's a sidebar with a red "COMPOSE" button, an "Inbox (7)" count, and sections for Starred, Drafts, and Sent Mail. Below that is a "Search people..." field and a list of contacts with their status (e.g., online, away, do not disturb) and names.

- spam / not spam
- priority level
- category (primary / social / promotions / updates)

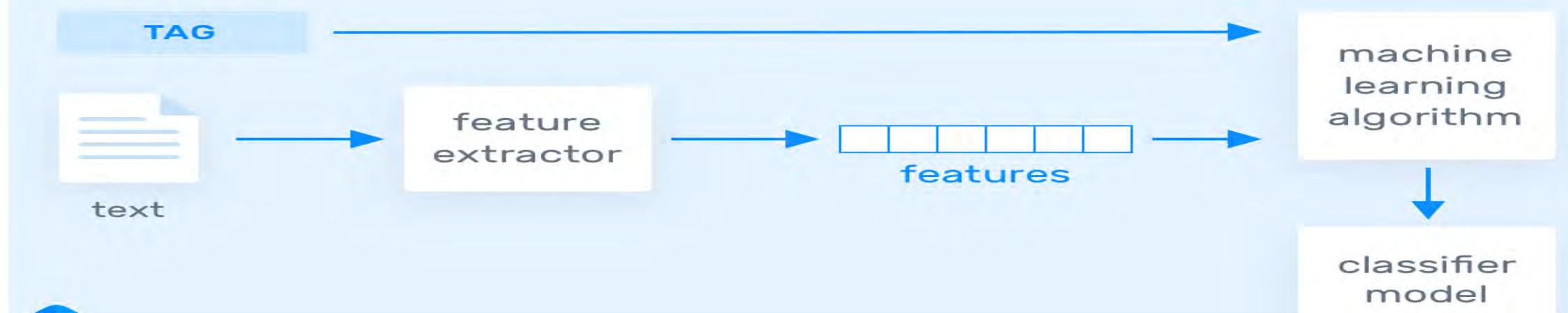
Applications of NLP: Text Classification

- question classification:

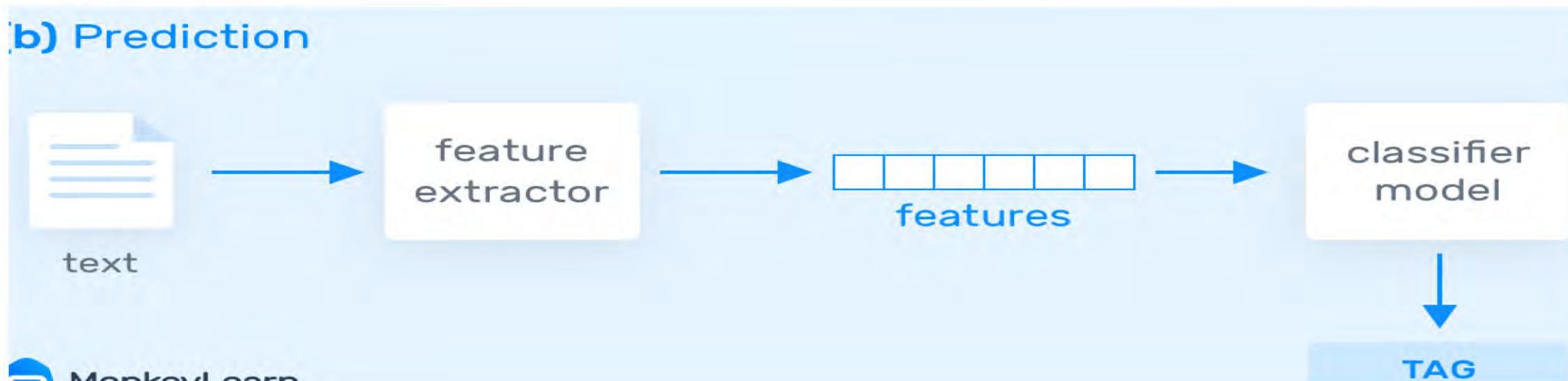
Who invented baseball ?	human
CNN is an acronym for what ?	abbreviation
Which Latin American country is the largest ?	location
How many small businesses are there in the U.S .	number
What would you add to the clay mixture to produce bone china ?	entity
What is the root of all evil ?	description

Applications of NLP: How Does Text Classification Work?

(a) Training



(b) Prediction



Applications of NLP: Speech Recognition

- Speech recognition technology uses natural language processing to transform spoken language into a machine-readable format.
- Speech recognition systems are an essential part of virtual assistants, like Siri, Alexa, and Google Assistant, for example. However, there are more and more use cases of speech recognition in business. For example, adding speech-to-text capabilities to business software, companies are able to automatically transcribe calls, send emails, and even translate.
- There are a lot of fields where speech recognition is used like, virtual assistants, adding speech-to-text, translating speech, sending emails etc.
- It is used in search engines where the user can voice out the name of their search requirements and get the desired result, making our work easier than typing out the entire command.

Applications of NLP: Chatbots & Virtual Assistants

- Voice assistant is a software that uses NLP and speech recognition to understand voice commands of a user and perform accordingly.
- Similarly, Chatbots are programs that are designed to assist an user 24/7 and respond appropriately and answer any query that the user might have.
- Most Chatbots and Virtual Assistants have pre-programmed answering systems that follow specific rules and patterns while answering.
- Powerful AI has enabled some voice assistants to interact with the user and respond appropriately. With more usage, they even improve themselves.
- Assistants like Siri and Alexa can even have a conversation with the user like a normal human being.
- Voice assistants like Alexa, Siri and Google Assistant, and chatbots that are integrated in many websites to help and guide new users.

Applications of NLP: Grammar Checkers and Auto Prediction

- Grammar Checking tools like Grammarly provides tons of features that help a person in writing better content. They can change any ordinary piece of text into beautiful literature.
- These softwares offer a lot of features like suggesting synonyms, correcting grammar and spellings, rephrasing sentences and giving clarity to the document and can even predict the tone of the sentence that might be implied by the user.
- Auto prediction is also a feature developed through NLP where the computer suggests automatic prediction of the text we have started typing. This saves time of the user and makes the job easier for them.

Applications of NLP: Grammar Checkers and Auto Prediction

The most common type of marketing channel is the wholesale market.

Varies kinds of **produce** are supplied from different areas are assembled at one place
and sold thru vegetables s
Replace the word
products
naller regional markets, etc. Fruits and
market handling and transport methods.

 Dismiss

Suggested by Grammarly



Applications of NLP: Advertisement to Targeted Audience

- If you ever search any product or object in any shopping site, you would often see ads of those products and other related products on other sites. This type of targeted online advertising is done with the help of NLP and it is known as Targeted Advertising.
- Through NLP, keywords that are searched by the user are matched with the keywords of the product ad. If they are similar, the user gets an advertisement. This process is called keyword matching.
- This has been highly beneficial to many companies and saved them a lot of investment as the ads are only shown to customers who are actually interested in the product, which is determined on the basis of their online activity.

Applications of NLP: Information Extraction

- Information extraction, automatically detects specific information in a text, such as names, companies, places, and more. This is also known as named entity recognition.
- You can also extract keywords within a text, as well as pre-defined features such as product serial numbers and models.

Test with your own text

Elon Musk has shared a photo of the spacesuit designed by SpaceX. This is the second image shared of the new design and the first to feature the spacesuit's full-body look.

Extract Text

Results

TAG	VALUE
KEYWORD	spacesuit
KEYWORD	Elon Musk
KEYWORD	full-body look
KEYWORD	second image
KEYWORD	new design
KEYWORD	photo

Applications of NLP: Information Extraction

- For example, many Wikipedia pages have an “infobox” that provides structured information about an entity or event.

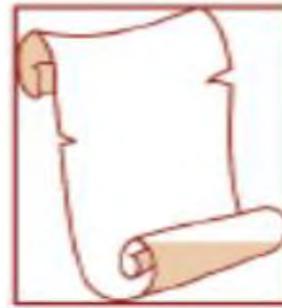
Studio album by Neutral Milk Hotel	
Released	February 10, 1998
Recorded	July–September 1997
Studio	Pet Sounds Studio, Denver, Colorado
Genre	Indie rock • psychedelic folk • lo-fi
Length	39:55
Label	Merge • Domino
Producer	Robert Schneider

(a) A Wikipedia infobox

- (17.1) In the Aeroplane Over the Sea is the second and final studio album by the American indie rock band Neutral Milk Hotel.
- (17.2) It was released in the United States on February 10, 1998 on Merge Records and May 1998 on Blue Rose Records in the United Kingdom.
- (17.3) Jeff Mangum moved from Athens, Georgia to Denver, Colorado to prepare the bulk of the album's material with producer Robert Schneider, this time at Schneider's newly created Pet Sounds Studio at the home of Jim McIntyre.
- (b) The first few sentences of text. Strings that match fields or field names in the infobox are underlined; strings that mention other entities are wavy underlined.

Figure 17.1: From the Wikipedia page for the album “In the Aeroplane Over the Sea”, retrieved October 26, 2017.

Applications of NLP: Information Extraction



President Barack Obama received the Serve America Act after congress' vote. He signed the bill last Thursday. The president said it would greatly increase service opportunities for the American people.

Applications of NLP: Machine Translation

- Machine translation (MT), process of translating one source language or text into another language, is one of the most important applications of NLP.
- Machine Translation has made a lot of improvement in the field due to availability of huge amounts of data and powerful machines, and advancement in the field of Machine learning and Neural networking.
- Even though Facebook's translations have been declared superhuman, machine translation still faces the challenge of understanding context.
- Automated translation is particularly useful in business because it facilitates communication, allows companies to reach broader audiences, and understand foreign documentation in a fast and cost-effective way.

Applications of NLP: Machine Translation

Nepali ▾ ↔ English ▾

स्कटल्याण्डको
ग्लास्गोमा आयोजित
सम्मेलनमा सहभागी
भएर प्रधानमन्त्री देउवा
आज बिहान काठमाडौं
आइपुगेका हुन् ।

Skaṭalyāṇḍakō glāsgōmā
āyōjita sam'mēlanamā
sahabhāgī bha'ēra

×

Prime Minister
Deuba arrived in
Kathmandu this
morning after
attending a
conference in
Glasgow, Scotland.

Applications of NLP: Text Summarization

- Text summarization is most helpfully applied in academic, research, or healthcare settings, as it uses natural language processing to quickly process text and extract the most important information.
- NLP can summarize text based on exact key phrases within the text, or it can even summarize based on determined meanings and inferences, providing a paraphrased summary.
- Its main goal is to simplify the process of going through vast amounts of data, such as scientific papers, news content, or legal documentation.
- There are two ways of using natural language processing to summarize data:
extraction-based summarization - which extracts key phrases and creates a summary, without adding any extra information - and **abstraction-based summarization**, which creates new phrases paraphrasing the original source. This second approach is more common and performs better.

Applications of NLP: Urgency Detection

- NLP techniques can also help you detect urgency in text.
- You can train an urgency detection model using your own criteria, so it can recognize certain words and expressions that denote gravity or discontent.
- This can help you prioritize the most important requests and make sure they don't get buried under a pile of unresolved tickets.
- NLP algorithms can be established that look for key phrases or words that connotate urgency or stress in text. This can help companies prioritize their work or customer service outreach to those who have communicated in a such a manner.
- Urgency detection helps you improve response times and efficiency, leading to a positive impact on customer satisfaction.

Applications of NLP: Question-answering

- Question-answering is a Computer Science discipline within the fields of AI and NLP. It focuses on building systems that automatically answer questions posted by human beings in their natural language.
- A computer system that understands the natural language has the capability of a program system to translate the sentences written by humans into an internal representation so that the valid answers can be generated by the system. The exact answers can be generated by doing syntax and semantic analysis of the questions.
- Question Answering is the task of answering questions (typically reading comprehension questions), but abstaining when presented with a question that cannot be answered based on the provided context.
- Popular benchmark datasets for evaluation question answering systems include SQuAD, HotPotQA, bAbI, TriviaQA, WikiQA, and many others.

Applications of NLP: Question-answering

Passage Sentence

In meteorology, precipitation is any product of the condensation of atmospheric water vapor that falls under gravity.

Question

What causes precipitation to fall?

Answer Candidate

gravity

- **Question Answering:**
 - More than search
 - Ask general comprehension questions of a document collection
 - Can be really easy: "What's the capital of Wyoming?"
 - Can be harder: "How many US states' capitals are also their largest cities?"
 - Can be open ended: "What are the main issues in the global warming debate?"
- **SOTA: Can do factoids, even when text isn't a perfect match**

The screenshot shows a Google search results page. The search bar at the top contains the query: "any US states' capitals are also their largest cities?". Below the search bar, there is a "Web" tab selected. The main search result area displays the following text:
Your search - How many US states' capitals are also their largest cities? - did not match any documents.
Suggestions:
- Make sure all words are spelled correctly.
- Try different keywords.
- Try more general keywords.
- Try fewer keywords.

At the bottom of the page, there are links to "Google Home", "Business Solutions", and "About Google". Below these links, there are two search results:
capital of Wyoming: Information From Answers.com
Note: click on a word meaning below to see its connections and related words.
The noun capital of Wyoming has one meaning: Meaning #1 : the capital.
www.answers.com/topic/capital-of-wyoming - 21k - [Cached](#) - [Similar pages](#)

Cheyenne: Weather and Much More From Answers.com
Cheyenne (shē-ĕn' , -ĕn') The capital of Wyoming, in the southeast part of the state near the Nebraska and Colorado borders.
www.answers.com/topic/cheyenne-wyoming - 74k - [Cached](#) - [Similar pages](#)

Applications of NLP: Text Generation

- Text generation is a subfield of natural language processing (NLP). It leverages knowledge in computational linguistics and artificial intelligence to automatically generate natural language texts, which can satisfy certain communicative requirements.
- Text generation three main scenarios:
 - **data-to text**, in which text is generated to explain or describe a structured record or unstructured perceptual input;
 - **text-to-text**, which typically involves fusing information from multiple linguistic sources into a single coherent summary; **and**
 - **dialogue**, in which text is generated as part of an interactive conversation with one or more human participants.

Applications of NLP: Data-to-text generation

- In data-to-text generation, the input ranges from structured records, such as the description of an weather forecast (as shown in Figure 19.1), to unstructured perceptual data, such as a raw image or video; the output may be a single sentence, such as an image caption, or a multi-paragraph argument. Despite this diversity of conditions, all data-to-text systems share some of the same challenges (Reiter and Dale, 2000):
 - determining what parts of the data to describe;
 - planning a presentation of this information;
 - lexicalizing the data into words and phrases;
 - organizing words and phrases into well-formed sentences and paragraphs.

Applications of NLP: Data-to-text generation

Temperature			
<i>time</i>	<i>min</i>	<i>mean</i>	<i>max</i>
06:00-21:00	9	15	21

Cloud sky cover			
<i>time</i>	<i>percent (%)</i>		
06:00-09:00	25-50		
09:00-12:00	50-75		

Wind speed			
<i>time</i>	<i>min</i>	<i>mean</i>	<i>max</i>
06:00-21:00	15	20	30

Wind direction			
<i>time</i>	<i>mode</i>		
06:00-21:00	S		

Cloudy, with temperatures between 10 and 20 degrees. South wind around 20 mph.

Figure 19.1: An example input-output pair for the task of generating text descriptions of weather forecasts (adapted from Konstas and Lapata, 2013).

Applications of NLP: Text-to-text generation

- Text-to-text generation includes problems of summarization and simplification:
 - Reading a novel and outputting a paragraph-long summary of the plot;
 - Reading a set of blog posts about politics, and outputting a bullet list of the various issues and perspectives;
 - Reading a technical research article about the long-term health consequences of drinking kombucha, and outputting a summary of the article in language that non-experts can understand

Applications of NLP: Dialog System

Dialog Systems

user: Schedule a meeting with Matt and David on Thursday.
computer: Thursday won't work for David. How about Friday?
user: I'd prefer Monday then, but Friday would be ok if necessary.

- ***NLP is hard***
 - Human languages are messy, ambiguous, and ever-changing
- ***What challenges get in the way of understanding and responding to natural language?***
 - Implicit references
 - Ambiguous references / semantics
 - Imprecise rules
 - Myriad languages
 - Scale
- ***Ambiguity and Variability of linguistic expression:***
 - Ambiguity: one form can mean many things
 - Variability: many forms can mean the same thing

NLP Challenges: Scale

- Examples:

- Bible (King James version): ~700K
- Penn Tree bank ~1M from Wall street journal
- Newswire collection:500M+
- Wikipedia: 2.9 billion word (English)
- Web: several billions of words

NLP Challenges: Language is dynamic

LOL	Laugh out loud
G2G	Got to go
BFN	Bye for now
B4N	Bye for now
Idk	I don't know
FWIW	For what it's worth
LUWAMH	Love you with all my heart

NLP Challenges: Language is Compositional



小心:
Carefully
Careful
Take
Care
Caution



地滑:
Slide
Landslip
Wet Floor
Smooth

Translate

English Spanish French Chinese - detected English Spanish Arabic Translate

小心地滑 Carefully slide

Xǐngxīn dì huá

NLP Challenges: Ambiguity

- Ambiguity in NLP refers to sentences and phrases that potentially have two or more possible interpretations.
 - ***Lexical ambiguity:*** a word that could be used as a verb, noun, or adjective.
 - ***Semantic ambiguity:*** the interpretation of a sentence in context. For example: I saw the boy on the beach with my binoculars. This could mean that I saw a boy through my binoculars or the boy had my binoculars with him.
 - ***Syntactic ambiguity:*** In the sentence above, this is what creates the confusion of meaning. The phrase with my binoculars could modify the verb, “saw,” or the noun, “boy.”
- Even for humans this sentence alone is difficult to interpret without the context of surrounding text. POS (part of speech) tagging is one NLP solution that can help solve the problem, somewhat.

NLP Challenges: Word Sense Ambiguity

- Many words have multiple meanings
- ***Homonymy***: the multiple meanings are unrelated
 - bank*₁: financial institution, *bank*₂: sloping land
 - bat*₁: club for hitting a ball, *bat*₂: nocturnal flying mammal
- ***Homographs***: same spelling, different meanings
 - Bank/Bank, Bat/Bat
- ***Homophones***: same pronunciation, different meanings
 - Write/Right, Piece/Peace

NLP Challenges: Word Sense Ambiguity

- Homonymy causes problems for NLP
 - information retrieval
 - query: *bat care*
 - machine translation
 - *bat*: *murciélagos* (animal) or *bate* (for baseball)
 - text-to-speech
 - *bass* (stringed instrument) vs. *bass* (fish)

NLP Challenges: Colloquialisms and slang

- Informal phrases, expressions, idioms, and culture-specific lingo present a number of problems for NLP – especially for models intended for broad use.
- Because as formal language, colloquialisms may have no “dictionary definition” at all, and these expressions may even have different meanings in different geographic areas.
- Furthermore, cultural slang is constantly morphing and expanding, so new words pop up every day.

NLP Challenges: Domain-specific language

- Different businesses and industries often use very different language.
- An NLP processing model needed for healthcare, for example, would be very different than one used to process legal documents.
- These days, however, there are a number of analysis tools trained for specific fields, but extremely niche industries may need to build or train their own models.

NLP Challenges: Phrases with multiple intentions

- Some phrases and questions actually have multiple intentions, so your NLP system can't oversimplify the situation by interpreting only one of those intentions.
- “*You’re late*”. What’s the speaker’s intention: informing or criticizing?

- The same expression means different things in different context.
 - *Where's the water? (Chemistry lab? Must be pure)*
 - *Where's the water? (Thirsty? Must be drinking water)*
 - *Where's the water? (Leaky roof? It can be dirty)*
- No natural language program can be complete because of new words, expression, and meaning can be generated quite freely.
 - *I'll fax it to you*
- There are lots of ways to say the same thing.
 - *Ram was born on October 11.*
 - *Ram's birthday is October 11.*

- Sentence and phrases might have hidden meanings
 - “*Out of sight, out of mind*”-> “ *invisible idiot*”
 - “*The spirit was willing but the flesh was weak*” - > “ *the vodka was good, but the meat was bad*”
- Problem due to extensive use of pronouns. (semantic issue)
 - *E.g. Ravi went to the supermarket. He found his favorite brand of coffee in rack. He paid for it and left.*
 - *It denotes??*
- Use of conjunctions to avoid repetition of phrases cause problem in NLP
 - *E.g. Ram and Hari went to restaurant. While Ram had a cup of coffee, Hari had tea. Hari had a cup of tea.*

Thank You

???

- **References:**

- 1) Foundations of Statistical Natural Language Processing - Christopher D. Manning.
- 2) Text Data Management and Analysis - A Practical Introduction to Information Retrieval and Text Mining
- 3) Natural Language Processing using Python- Apress (2019)
- 4) Applied Text Analysis with Python O'Reilly Media (2018)
- 5) Natural Language Processing Python and NLTK-Packt Publishing (2016)
- 6) Building Machine Learning Systems with Python-Packt Publishing (2015)