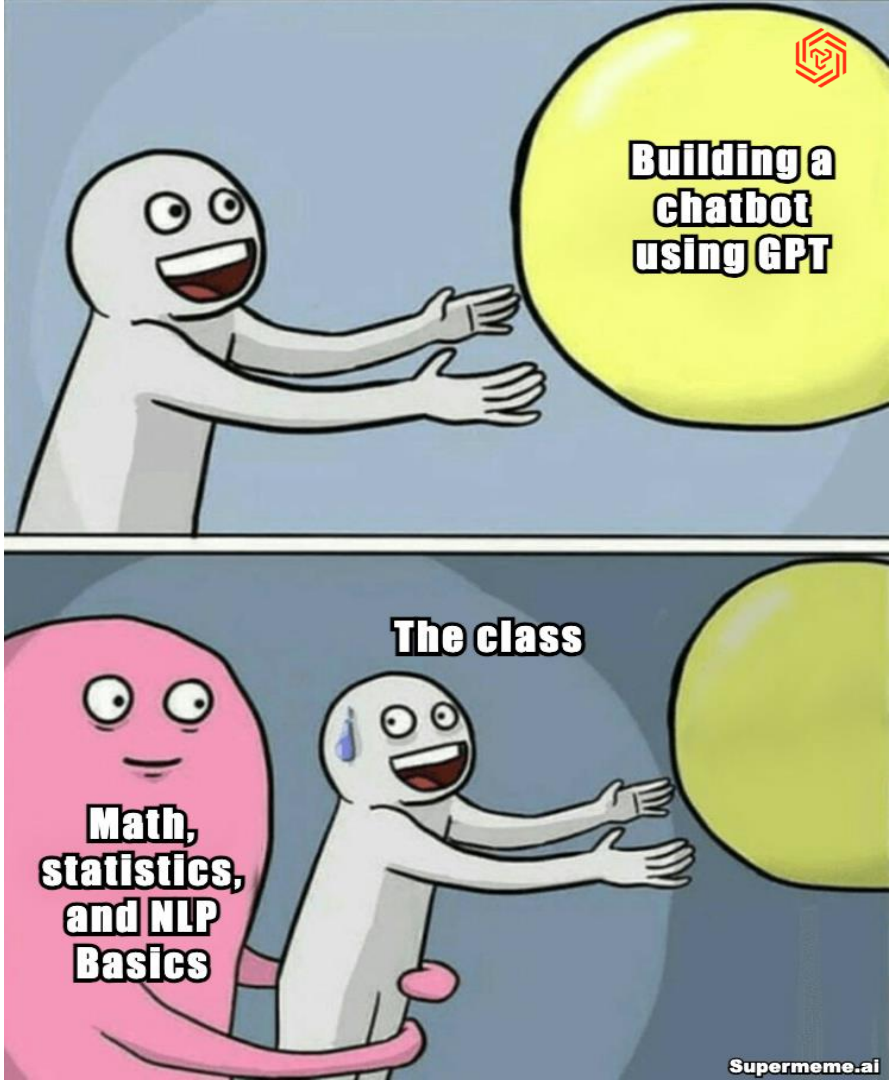
# Building Large Language Model Applications

## Introduction to Natural Language Processing

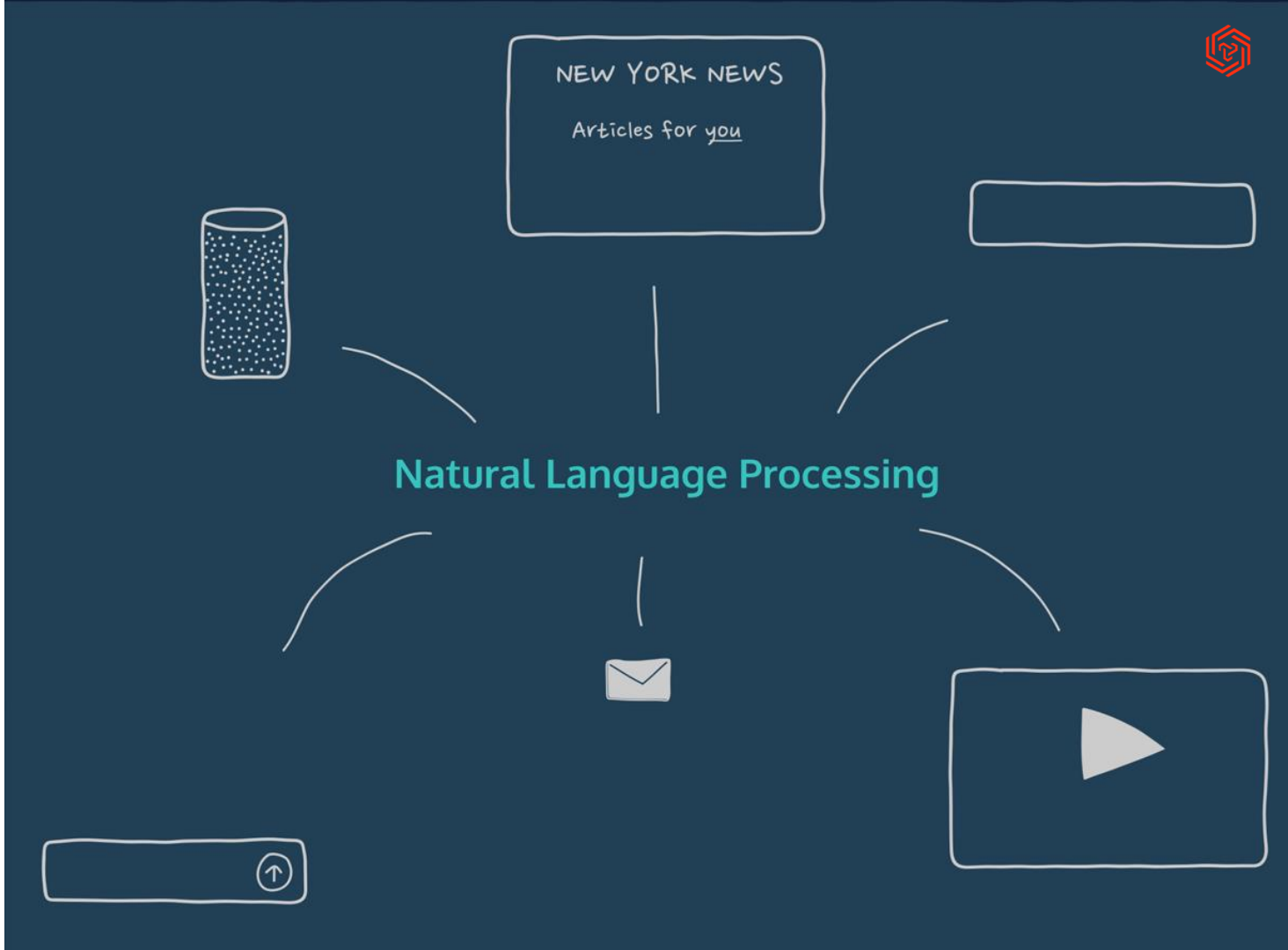
Hamza Farooq
Dr. Saima Hassan



# Learning outcomes

- NLP Overview
- History of NLP
- Applications of NLP
- Common NLP Tasks
- NLP Ambiguities
- Conclusion

# We live in a world of NLP



# What is NLP from the experts' perspective?

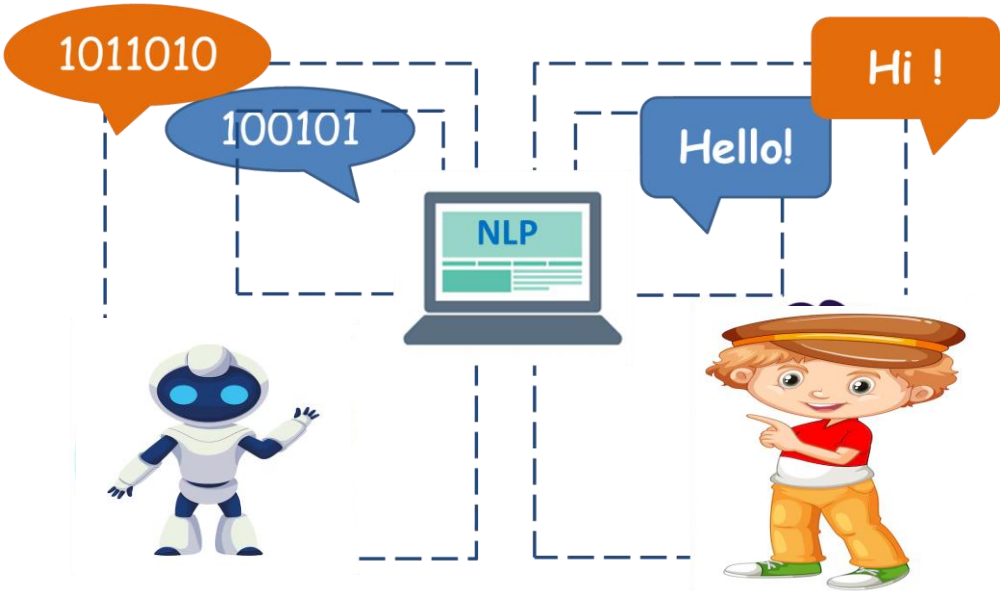
"Natural language is the most important part of artificial intelligence." **John Searle**

"Natural language processing is a cornerstone of artificial intelligence, allowing computers to read and understand human language, as well as to produce and recognize speech." **Ginni Rometty**

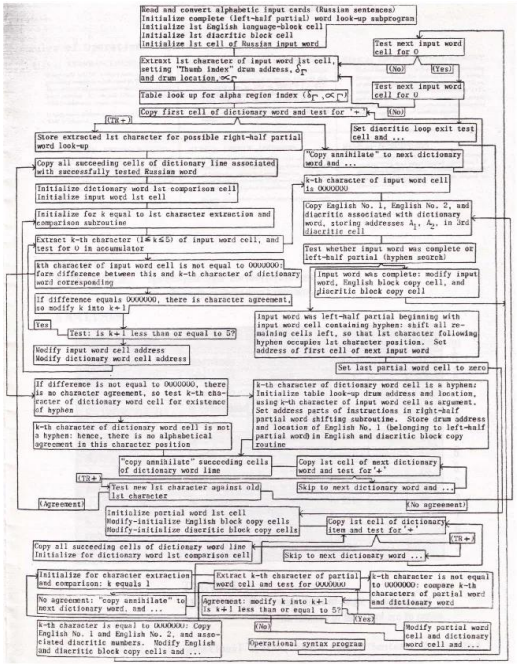

"Natural language processing is one of the most important fields in artificial intelligence and also one of the most difficult." **Dan Jurafsky**

# What is NLP?

- **NLP process information** contained in a **natural language text**
- It deals with the interaction between **human** and **computer** using **natural language**



# History of NLP?



A FLOWCHART OF PART OF THE IBM'S DICTIONARY LOOKUP PROCEDURES.

## 1950s

```
 Welcome to
              EEEEEE  LL      IIII   ZZZZZZ   AAAAA
              EE      LL       II       ZZ   AA   AA
              EEEEE   LL       II     ZZZ    AAAAAAA
              EE      LL       II    ZZ      AA   AA
              EEEEEE  LLLLLL  IIII ZZZZZZ    AA   AA

  Eliza is a mock Rogerian psychotherapist.
  The original program was described by Joseph Weizenbaum in 1966.
  This implementation by Norbert Landsteiner 2005.


ELIZA: Is something troubling you ?
YOU:   Men are all alike.
ELIZA: What is the connection, do you suppose ?
YOU:   They're always bugging us about something or other.
ELIZA: Can you think of a specific example ?
YOU:   Well, my boyfriend made me come here.
ELIZA: Is it important to you that your boyfriend made you come here ?
YOU:   He says I'm depressed much of the time.
ELIZA: I am sorry to hear that you are depressed.
YOU:   It's true. I am unhappy.
ELIZA: Can you explain what made you unhappy ?
YOU:
```

https://en.wikipedia.org/wiki/ELIZA

## 1960s

```
(defrule 52
 if (site culture is blood)
  (gram organism is neg)
  (morphology organism is rod)
  (burn patient is serious)
 then .4
  (identity organism is pseudomonas))
Rule 52:
 If
  1) THE SITE OF THE CULTURE IS BLOOD
  2) THE GRAM OF THE ORGANISM IS NEG
  3) THE MORPHOLOGY OF THE ORGANISM IS ROD
  4) THE BURN OF THE PATIENT IS SERIOUS
 Then there is weakly suggestive evidence (0.4) that
  1) THE IDENTITY OF THE ORGANISM IS PSEUDOMONAS
```

## 1970s and 80s

Gemini
1.5 Flash

write a poem on AI in four lines

Here's a poem on AI in four lines:

A mind of metal, thoughts take flight,
Learning fast, both day and night.
Solving problems, complex and grand,
AI's future, close at hand.

I hope this resonates with you! Let me know if you'd like a different style or theme.

## Today

# Go beyond the keyword matching



- Identify the **structure** and **meaning** of words, **sentences, texts** and **conversations**
- **Deep** understanding of **broad** language
- NLP is all around us

# Knowledge requirements for Machine

**Phonetics and Phonology:** Knowledge about linguistic sounds

**Morphology:** Knowledge of the meaningful components of words

**Syntax:** Knowledge of the structural relationships between words

**Semantics:** Knowledge of meaning

**Pragmatics:** Knowledge of the relationship of meaning to the goals and intentions of the speaker

**Discourse:** Knowledge about linguistic units larger than a single utterance

# Applications of NLP

# What are the NLP applications you interact with daily?

- Search engines
- Google Translate
- Social Media
- Job Seeking
- ...

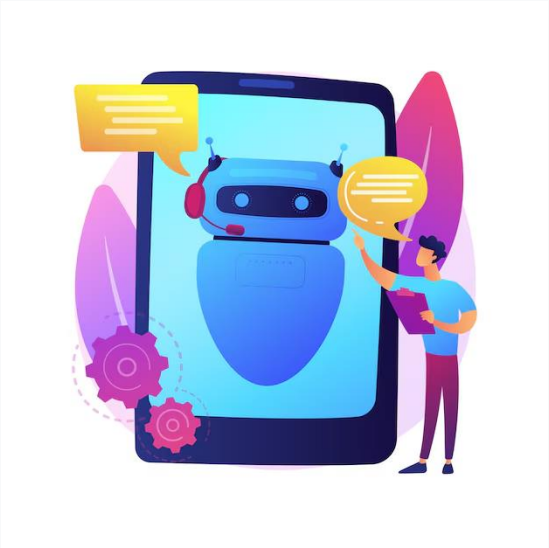


Most ideas stem from Academia, but big guys have (several) strong NLP research labs (like Microsoft, Yahoo, AT&T, IBM, etc.)

# Applications

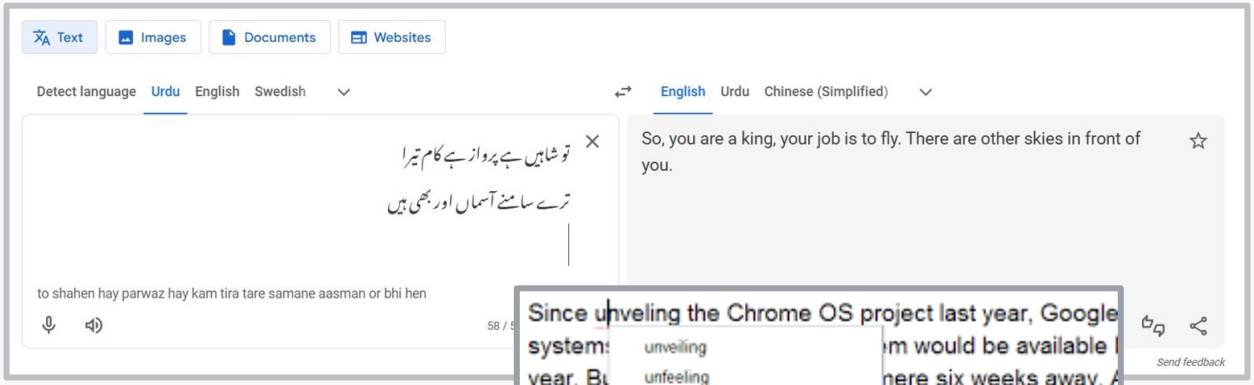
**Conversational agents contains:**

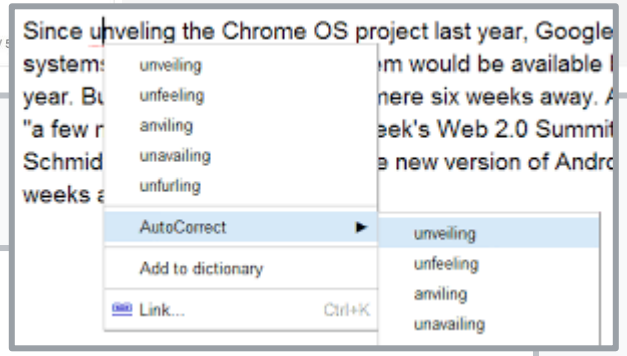
- Speech recognition
- Language analysis
- Dialogue processing
- Information retrieval
- Text to speech

# Applications

- Machine Translation
- Summarization
- Auto Completion
- Spell Correction



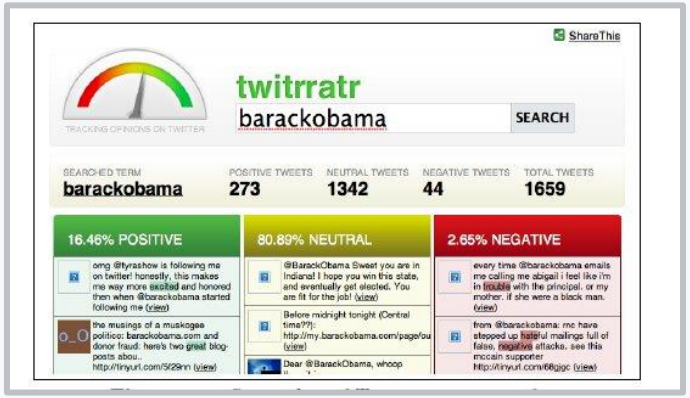

## PDF SUMMARY

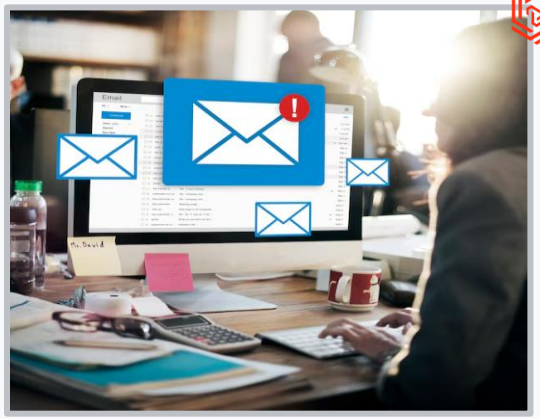
This study analyzes public sentiment towards AI tools using 500,000 tweets from January to March 2023. Several NLP models, including BERT, RoBERTa, DistilBERT, Sci-BERT, and VADER, were used for sentiment classification (positive, negative, neutral). RoBERTa achieved the highest accuracy (95.36%), while other models showed accuracy above 92%. The research provides insights into public opinion on AI tools during this period.

# Applications

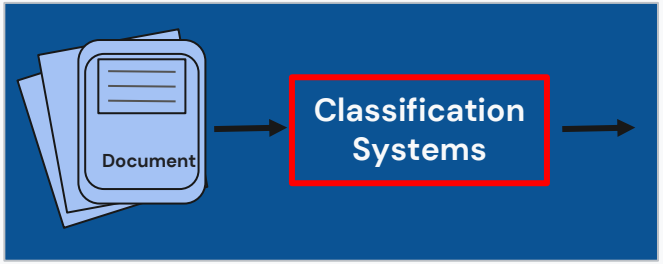
- Sentiment Analysis
- Text Classification

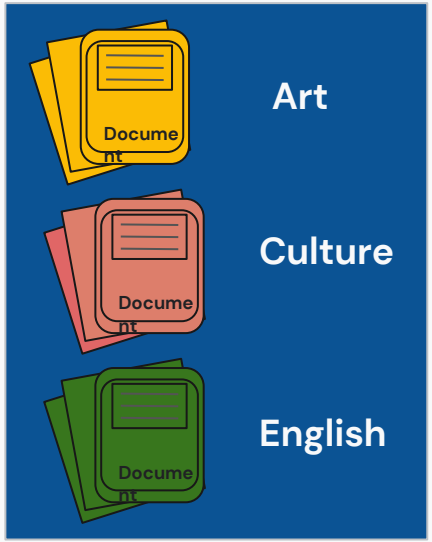
**Many More...**



DOI: 10.13140/RG.2.1.1809.1044



# Common NLP tasks

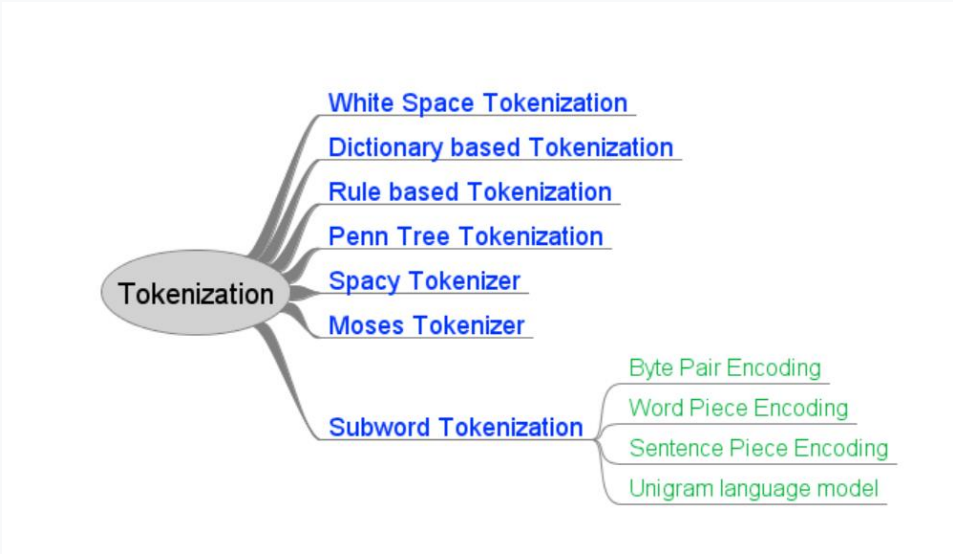
# Common NLP tasks

- **Tokenization**
- POS Tagging
- Word Sense Disambiguation
- Dependency Parsing
- Syntactic Parsing
- Semantic Analysis
- Coreference Resolution
- Named Entity Recognition (NER)
- Text Representation
- Text Classification
- Natural Language Understanding
- Natural Language Generation
- Natural Language Translation
- Multimodal NLP

- **Tokenization** is the process of breaking down a text into individual units called tokens.
- **Tokens** are typically **words**, but can also be phrases or even individual characters, depending on the application.
- Tokenization is a crucial step in natural language processing tasks such as **machine translation, sentiment analysis, and named entity recognition.**

# Common NLP tasks

- **Tokenization**
- POS Tagging
- Word Sense Disambiguation
- Dependency Parsing
- Syntactic Parsing
- Semantic Analysis
- Coreference Resolution
- Named Entity Recognition (NER)
- Text Representation
- Text Classification
- Natural Language Understanding
- Natural Language Generation
- Natural Language Translation
- Multimodal NLP

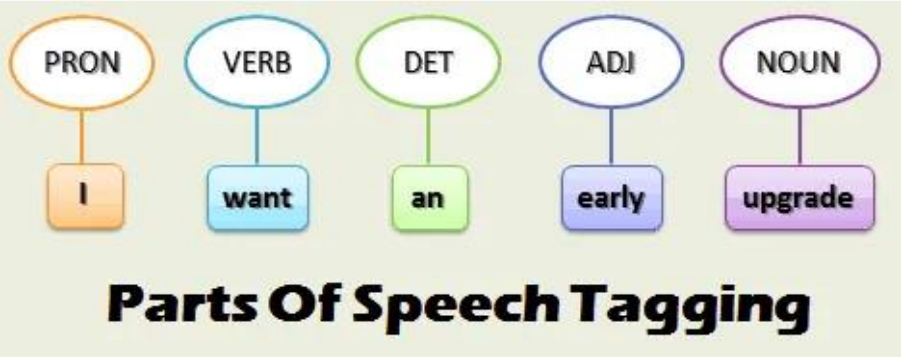
# Common NLP tasks

- Tokenization
- **POS Tagging**
- Word Sense Disambiguation
- Dependency Parsing
- Syntactic Parsing
- Semantic Analysis
- Coreference Resolution
- Named Entity Recognition (NER)
- Text Representation
- Text Classification
- Natural Language Understanding
- Natural Language Generation
- Natural Language Translation
- Multimodal NLP

- **POS** stands for **Part-of-Speech**, which is a linguistic term used to describe the **grammatical category of a word in a sentence.**
- **POS** tagging is the process of assigning each word in a text with its corresponding POS category, such as **noun, verb, adjective,** or **adverb**.
- **POS** tagging is a critical component in various natural language processing tasks, including **text-to-speech conversion, information retrieval,** and **machine translation**.

# Common NLP tasks

- Tokenization
- **POS Tagging**
- Word Sense Disambiguation
- Dependency Parsing
- Syntactic Parsing
- Semantic Analysis
- Coreference Resolution
- Named Entity Recognition (NER)
- Text Representation
- Text Classification
- Natural Language Understanding
- Natural Language Generation
- Natural Language Translation
- Multimodal NLP




**“John eats pizza.”**

POS tagging would label "**John**" as a **proper noun** and "**eats**" as a **verb**, while syntactic parsing would identify "**John**" as the **subject** of the verb "eats" and "**pizza**" as the **object** of the verb.

POS tagging is concerned with the individual words, while syntactic parsing focuses on the overall sentence structure.

# Common NLP tasks

- Tokenization
- POS Tagging
- **Word Sense Disambiguation**
- Dependency Parsing
- Syntactic Parsing
- Semantic Analysis
- Coreference Resolution
- Named Entity Recognition (NER)
- Text Representation
- Text Classification
- Natural Language Understanding
- Natural Language Generation
- Natural Language Translation
- Multimodal NLP

"The chicken is ready to eat."

Is the chicken cooked and ready for someone to eat, or is the chicken hungry and ready to eat?

# Common NLP tasks

- Tokenization
- POS Tagging
- **Word Sense Disambiguation**
- Dependency Parsing
- Syntactic Parsing
- Semantic Analysis
- Coreference Resolution
- Named Entity Recognition (NER)
- Text Representation
- Text Classification
- Natural Language Understanding
- Natural Language Generation
- Natural Language Translation
- Multimodal NLP

- Word sense disambiguation is the process of identifying the **correct meaning of a word** with **multiple possible meanings** based on the context in which it appears.
- This is a crucial task in natural language processing because words often have different meanings depending on the context in which they are used.
- Word sense disambiguation is used in various applications, including **information retrieval, machine translation, and question answering systems.**

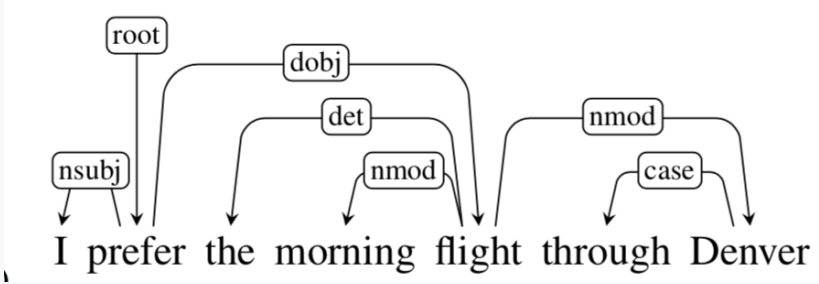
# Common NLP tasks

- Tokenization
- POS Tagging
- Word Sense Disambiguation
- **Dependency Parsing**
- Syntactic Parsing
- Semantic Analysis
- Coreference Resolution
- Named Entity Recognition (NER)
- Text Representation
- Text Classification
- Natural Language Understanding
- Natural Language Generation
- Natural Language Translation
- Multimodal NLP

- Dependency parsing is the process of analyzing the **grammatical structure of a sentence** by identifying the relationships between words in a sentence.
- It involves identifying the subject, object, and other dependent clauses and phrases, and representing them as a tree-like structure known as a dependency tree.
- Dependency parsing is used in various natural language processing applications, including **sentiment analysis, named entity recognition, and machine translation**.

# Common NLP tasks

- Tokenization
- POS Tagging
- Word Sense Disambiguation
- **Dependency Parsing**
- Syntactic Parsing
- Semantic Analysis
- Coreference Resolution
- Named Entity Recognition (NER)
- Text Representation
- Text Classification
- Natural Language Understanding
- Natural Language Generation
- Natural Language Translation
- Multimodal NLP




**Head-Dependent**: In the arrows representing relationship, the origin word is the Head & the destination word is Dependent.
**Root**: Word which is the root of our parse tree. It is 'prefer' in the above example.
**Grammar Functions and Arcs**: Tags between each Head-Dependent pair is a grammar function determining the relation between the Head & Dependent. The arrowhead carrying the tag is called an Arc.

| Clausal Argument Relations | Description |
|---|---|
| NSUBJ | Nominal subject |
| DOBJ | Direct object |
| IOBJ | Indirect object |
| CCOMP | Clausal complement |
| XCOMP | Open clausal complement |
| **Nominal Modifier Relations** | **Description** |
| NMOD | Nominal modifier |
| AMOD | Adjectival modifier |
| NUMMOD | Numeric modifier |
| APPOS | Appositional modifier |
| DET | Determiner |
| CASE | Prepositions, postpositions and other case markers |
| **Other Notable Relations** | **Description** |
| CONJ | Conjunct |
| CC | Coordinating conjunction |

# Common NLP tasks

- Tokenization
- POS Tagging
- Word Sense Disambiguation
- Dependency Parsing
- **Syntactic Parsing**
- Semantic Analysis
- Coreference Resolution
- Named Entity Recognition (NER)
- Text Representation
- Text Classification
- Natural Language Understanding
- Natural Language Generation
- Natural Language Translation
- Multimodal NLP

- Syntactic parsing is the process of **analyzing the grammatical structure** of a sentence to determine its syntactic components, such as **nouns, verbs, adjectives,** and **adverbs.**
- It involves identifying the parts of speech of each word in the sentence and grouping them together into phrases and clauses based on their syntactic relationships.
- Syntactic parsing is used in various natural language processing applications, including **text-to-speech conversion, machine translation,** and **information retrieval.**

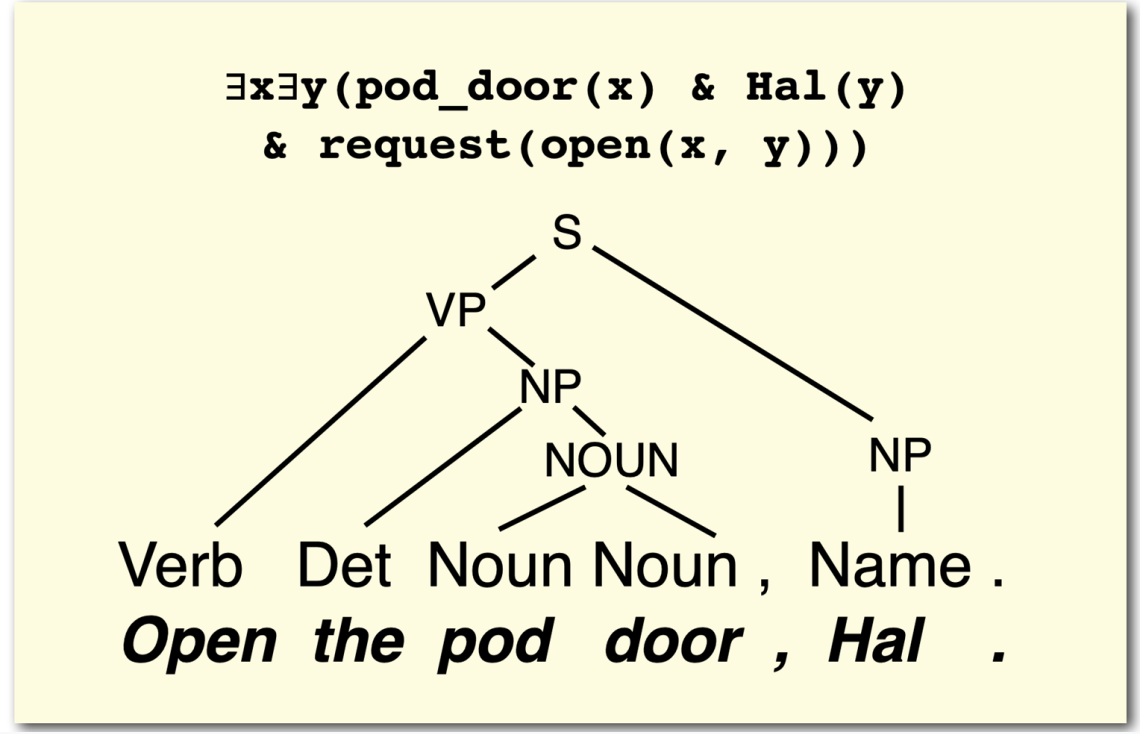
# Common NLP tasks

- Tokenization
- POS Tagging
- Word Sense Disambiguation
- Dependency Parsing
- **Syntactic Parsing**
- Semantic Analysis
- Coreference Resolution
- Named Entity Recognition (NER)
- Text Representation
- Text Classification
- Natural Language Understanding
- Natural Language Generation
- Natural Language Translation
- Multimodal NLP

- Semantic analysis is the process of **extracting the meaning of a text** by analyzing the relationships between words and phrases in a sentence.
- It involves identifying the underlying concepts and ideas conveyed by the text and representing them in a structured form, such as a **knowledge graph or ontology**.
- Semantic analysis is used in various natural language processing applications, including **question answering, information retrieval,** and **chatbots**, to enable more accurate and intelligent responses.

# Common NLP tasks

- Tokenization
- POS Tagging
- Word Sense Disambiguation
- Dependency Parsing
- Syntactic Parsing
- **Semantic Analysis**
- Coreference Resolution
- Named Entity Recognition (NER)
- Text Representation
- Text Classification
- Natural Language Understanding
- Natural Language Generation
- Natural Language Translation
- Multimodal NLP



# Common NLP tasks

- Tokenization
- POS Tagging
- Word Sense Disambiguation
- Dependency Parsing
- Syntactic Parsing
- **Semantic Analysis**
- Coreference Resolution
- Named Entity Recognition (NER)
- Text Representation
- Text Classification
- Natural Language Understanding
- Natural Language Generation
- Natural Language Translation
- Multimodal NLP

We need a meaning representation language.

**"Shallow" semantic analysis:** Template-filling (Information Extraction)

Named-Entity Extraction: Organizations, Locations, Dates,...
Event Extraction

**"Deep" semantic analysis:** (Variants of) formal logic

$$\exists x \exists y (\text{pod\_door}(x) \& \text{Hal}(y) \& \text{request}(\text{open}(x,y)))$$

We also distinguish between
**Lexical semantics** (the meaning of words) and
**Compositional semantics** (the meaning of sentences)

# Common NLP tasks

- Tokenization
- POS Tagging
- Word Sense Disambiguation
- Dependency Parsing
- Syntactic Parsing
- Semantic Analysis
- **Coreference Resolution**
- Named Entity Recognition (NER)
- Text Representation
- Text Classification
- Natural Language Understanding
- Natural Language Generation
- Natural Language Translation
- Multimodal NLP

More than a decade ago, Carl Lewis stood on the threshold of what was to become the greatest athletics career in history. He had just broken two of the legendary Jesse Owens' college records, but never believed he would become a corporate icon, the focus of hundreds of millions of dollars in advertising. His sport was still nominally amateur. Eighteen Olympic and World Championship gold medals and 21 world records later, Lewis has become the richest man in the history of track and field -- a multi-millionaire.

Who is Carl Lewis?
Did Carl Lewis break any world records?
(and how do you know that?)

# Common NLP tasks

- Tokenization
- POS Tagging
- Word Sense Disambiguation
- Dependency Parsing
- Syntactic Parsing
- Semantic Analysis
- **Coreference Resolution**
- Named Entity Recognition (NER)
- Text Representation
- Text Classification
- Natural Language Understanding
- Natural Language Generation
- Natural Language Translation
- Multimodal NLP

- Coreference resolution is the task of **identifying all the expressions** (e.g., pronouns, names) in a text that refer to the same entity, and linking them together.
- It is a crucial task in natural language processing as it enables a system to maintain a **consistent representation of entities throughout a document**, enabling more accurate information extraction and text understanding.

# Common NLP tasks

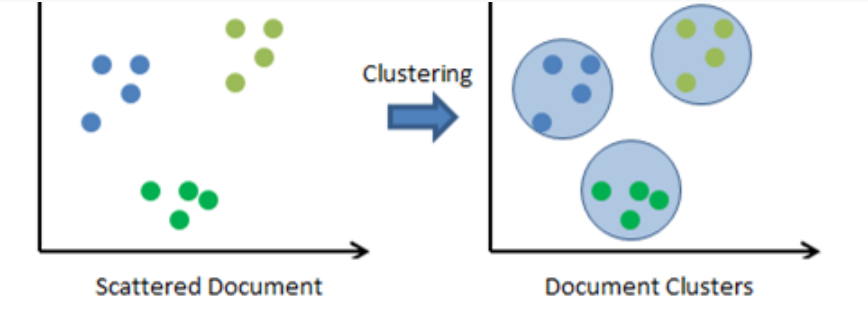
- Tokenization
- POS Tagging
- Word Sense Disambiguation
- Dependency Parsing
- Syntactic Parsing
- Semantic Analysis
- Coreference Resolution
- **Named Entity Recognition (NER)**
- Text Representation
- Text Classification
- Natural Language Understanding
- Natural Language Generation
- Natural Language Translation
- Multimodal NLP

Named entity recognition (NER) is the process of identifying and categorizing named entities in a text, such as people, organizations, locations, and dates.


*Figure 1:* An example of NER application on an example text

# Common NLP tasks

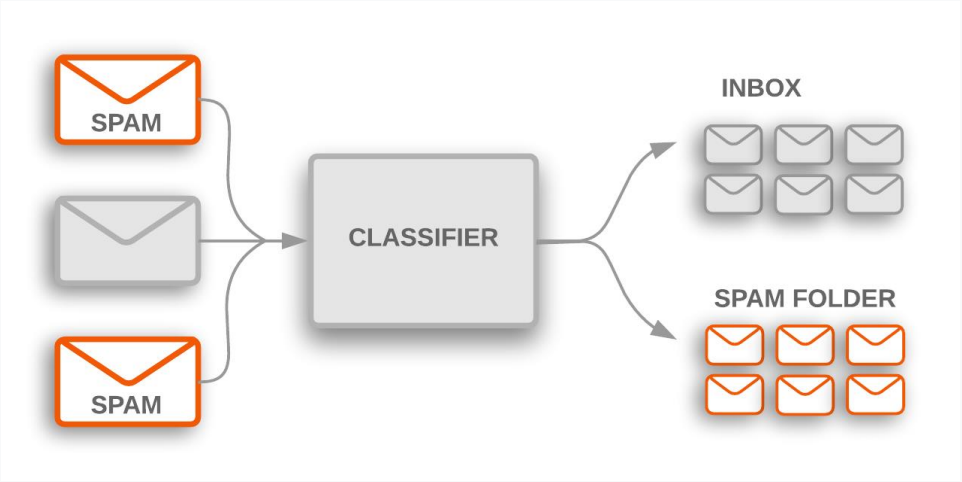
- Tokenization
- POS Tagging
- Word Sense Disambiguation
- Dependency Parsing
- Syntactic Parsing
- Semantic Analysis
- Coreference Resolution
- Named Entity Recognition (NER)
- **Text Representation**
- Text Classification
- Natural Language Understanding
- Natural Language Generation
- Natural Language Translation
- Multimodal NLP

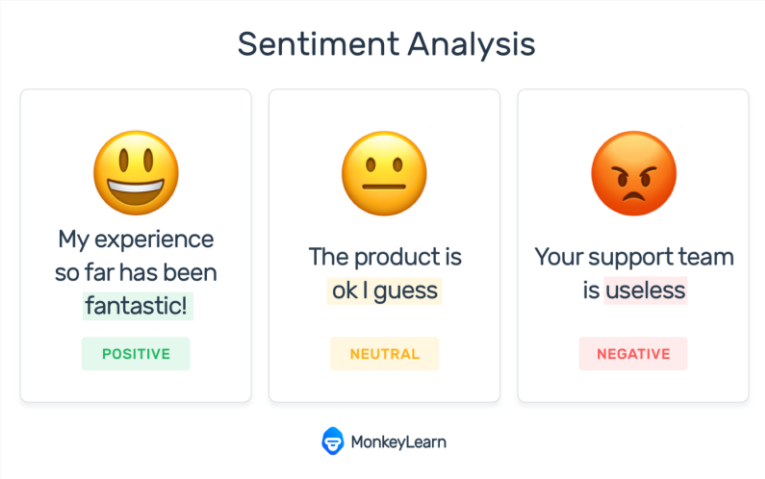



- Text representation is the process of converting **unstructured text data into a structured format** that can be used for natural language processing tasks.
- It involves selecting a suitable representation scheme, such as bag-of-words, word embeddings, or topic models, to capture the key features and characteristics of the text data in a numerical form that can be processed by machine learning algorithms.

# Common NLP tasks

- Tokenization
- POS Tagging
- Word Sense Disambiguation
- Dependency Parsing
- Syntactic Parsing
- Semantic Analysis
- Coreference Resolution
- Named Entity Recognition (NER)
- Text Representation
- **Text Classification**
- Natural Language Understanding
- Natural Language Generation
- Natural Language Translation
- Multimodal NLP



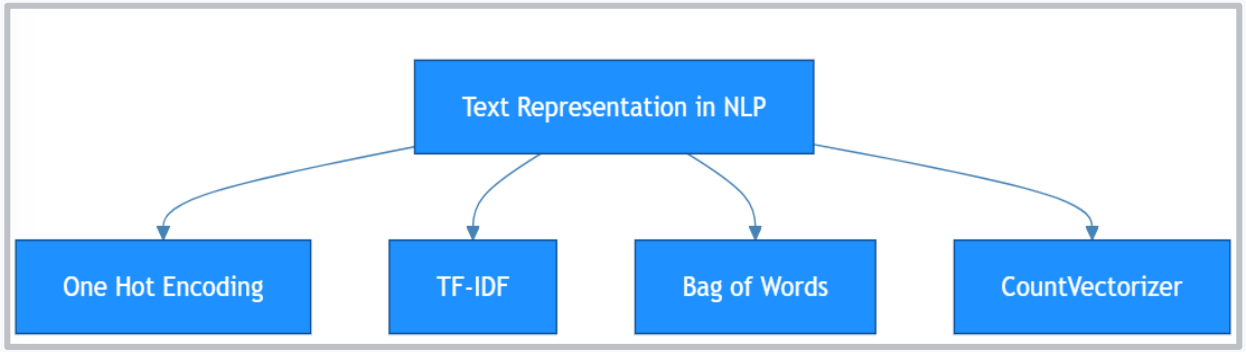



# Common NLP tasks

- Tokenization
- POS Tagging
- Word Sense Disambiguation
- Dependency Parsing
- Syntactic Parsing
- Semantic Analysis
- Coreference Resolution
- Named Entity Recognition (NER)
- Text Representation
- Text Classification
- **Natural Language Understanding**
- Natural Language Generation
- Natural Language Translation
- Multimodal NLP

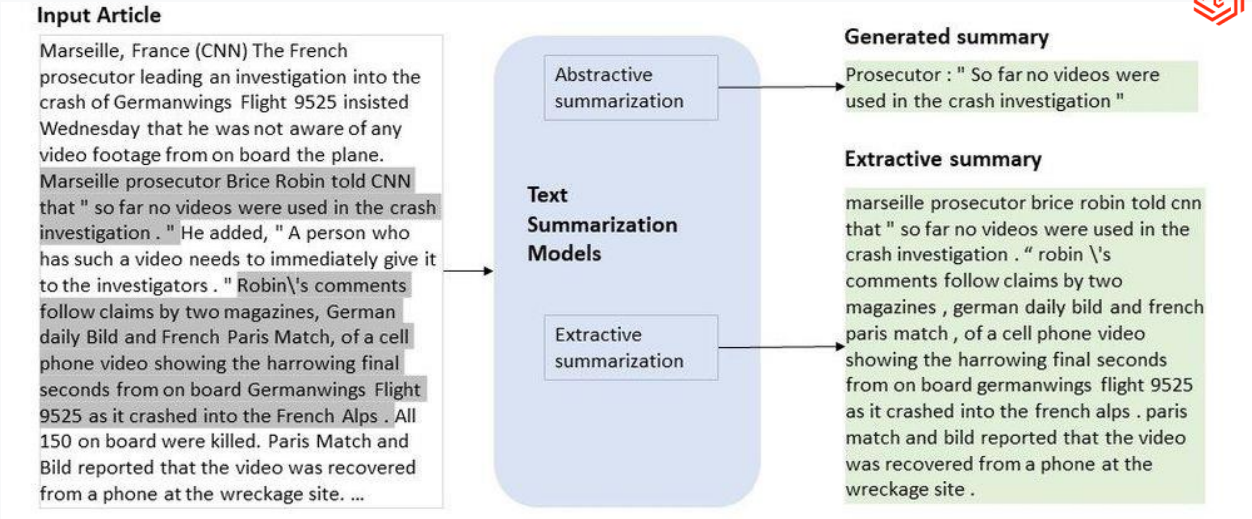
**Natural language understanding**

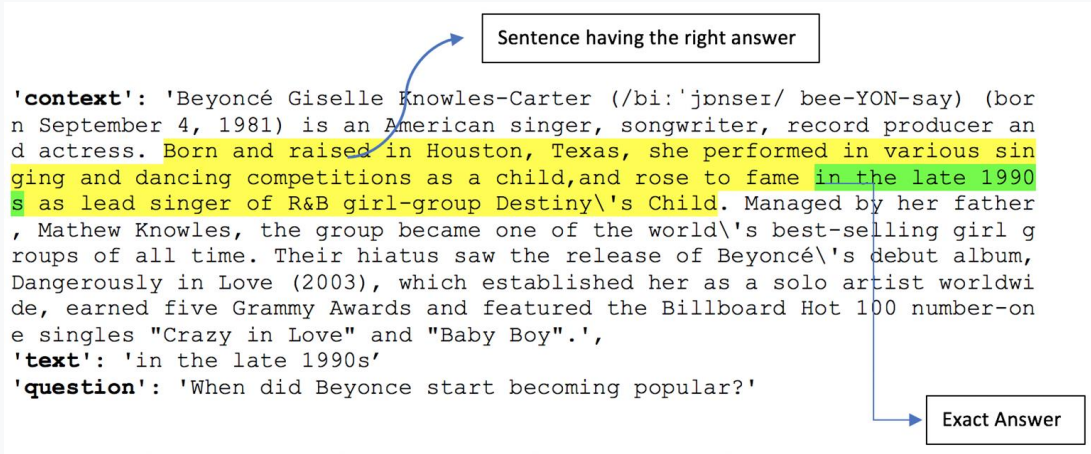
- Extract information (e.g. about entities or events) from text
- Translate raw text into a meaning representation
- Reason about information given in text
- Execute NL instructions



# Common NLP tasks

- Tokenization
- POS Tagging
- Word Sense Disambiguation
- Dependency Parsing
- Syntactic Parsing
- Semantic Analysis
- Coreference Resolution
- Named Entity Recognition (NER)
- Text Representation
- Text Classification
- Natural Language Understanding
- **Natural Language Generation**
- Natural Language Translation
- Multimodal NLP



# Common NLP tasks

- Tokenization
- POS Tagging
- Word Sense Disambiguation
- Dependency Parsing
- Syntactic Parsing
- Semantic Analysis
- Coreference Resolution
- Named Entity Recognition (NER)
- Text Representation
- Text Classification
- Natural Language Understanding
- Natural Language Generation
- **Natural Language Translation**
- Multimodal NLP

## Natural language translation

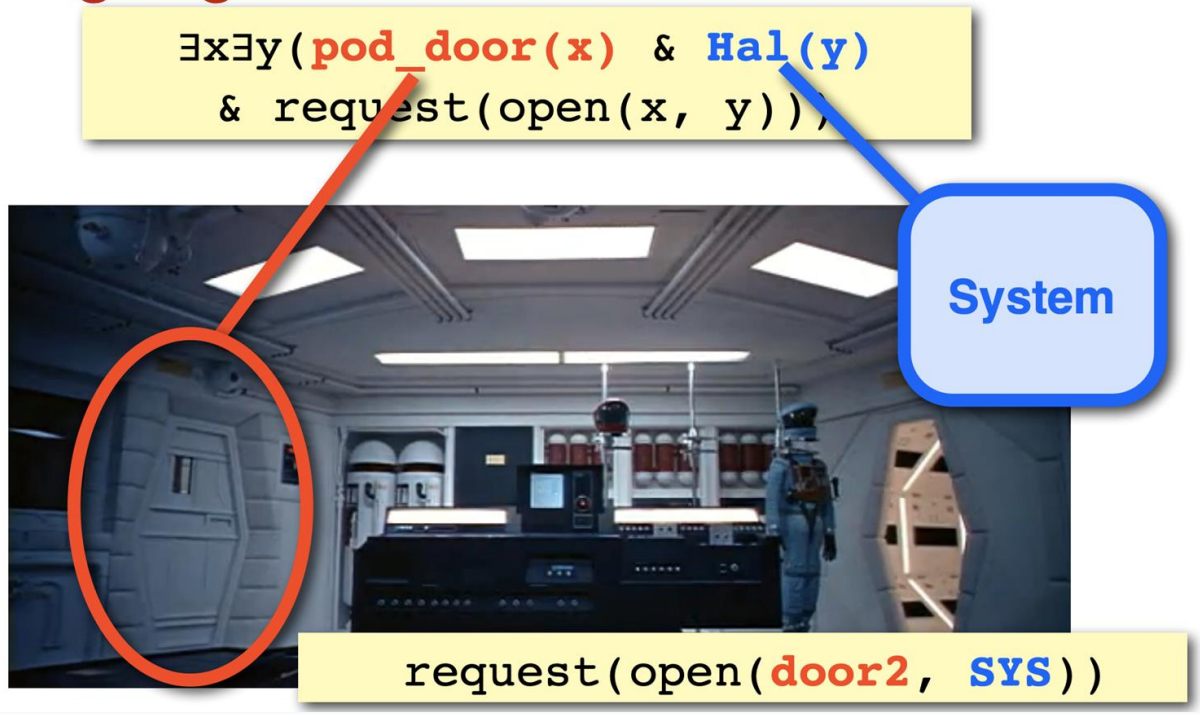
- Translate one natural language to another


https://www.shaip.com/blog/nlp-in-translation/

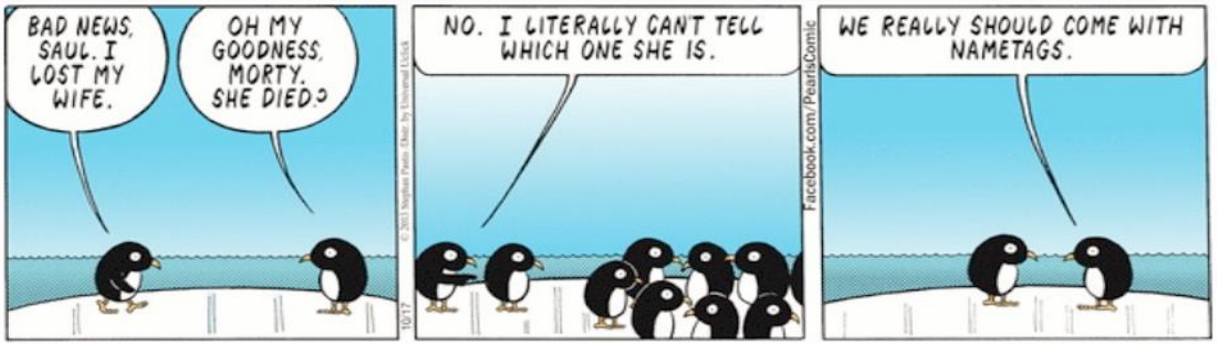
# Common NLP tasks

- Tokenization
- POS Tagging
- Word Sense Disambiguation
- Dependency Parsing
- Syntactic Parsing
- Semantic Analysis
- Coreference Resolution
- Named Entity Recognition (NER)
- Text Representation
- Text Classification
- Natural Language Understanding
- Natural Language Generation
- Natural Language Translation
- **Multimodal NLP**

## Multimodal NLP: mapping from language to the world



# NLP Ambiguities




credit: A. Zwicky

# NLP Ambiguities

**1. Lexical Ambiguity:** It is defined as the ambiguity associated with the meaning of a single word. A single word can have different meanings. Also, a single word can be a noun, adjective, or verb. For example, The word “bank” can have different meanings. It can be a financial bank or a riverbank. Similarly, the word “clean” can be a noun, adverb, adjective, or verb.


# NLP Ambiguities

**2. Syntactic Ambiguity**: It is defined as the ambiguity associated with the way the words are parsed. For example, The sentence "Visiting relatives can be boring." This sentence can have two different meanings. One is that visiting a relative's house can be boring. The second is that visiting relatives at your place can be boring.

# NLP Ambiguities

**3. Semantic Ambiguity**: This occurs when the meaning of words in a sentence can be interpreted in multiple ways. For instance, consider the sentence, *"Mary knows a little French."* Here, the phrase *"a little French"* is ambiguous because it could refer either to the French language or to a person of French origin.

# Conclusion

What is NLP?

Key challenges and applications

Common NLP tasks

NLP Ambiguities