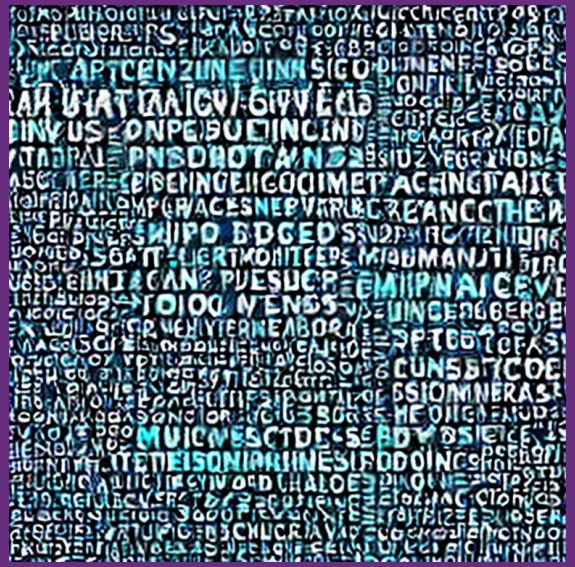
Mila

20-06-2024



Introduction to Natural Language Processing

Part 1

Prakhar Ganesh

Before we start ...

How’s everyone doing?

Any questions from previous sessions?

Goals today...

- What is '*Natural Language Processing (NLP)*'?
  - Introduction to an interdisciplinary field
- Why do we need NLP?
  - Applications and Challenges
- Different ways of modeling language
  - Bag of Words, Causal Language Modeling, etc.
- NLP Pre-Deep Learning
- Moving Towards NLP Post-Deep Learning: Embeddings

Some slides borrowed from lectures by Mélisande and Niki

What is Natural Language Processing?

What is Natural Language Processing?

It enables computers to understand, interpret and respond to human language.

What is Natural Language Processing?

*It enables **computers** to understand, interpret and respond to human language.*

- Computer Science, Artificial intelligence, Machine learning

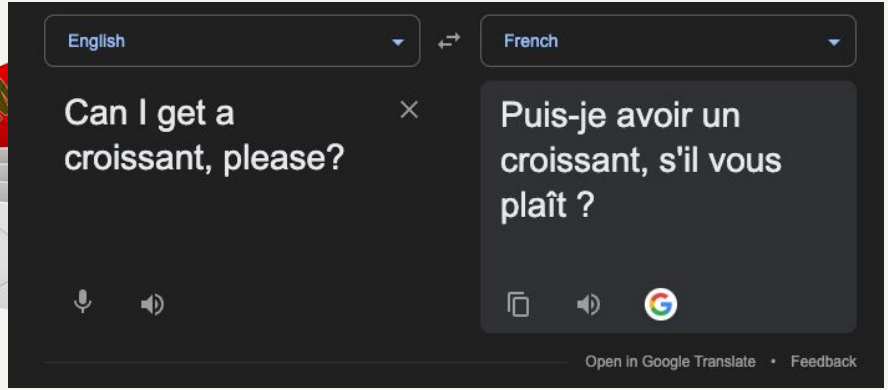
What is Natural Language Processing?

It enables computers to understand, interpret and respond to ***human language****.*

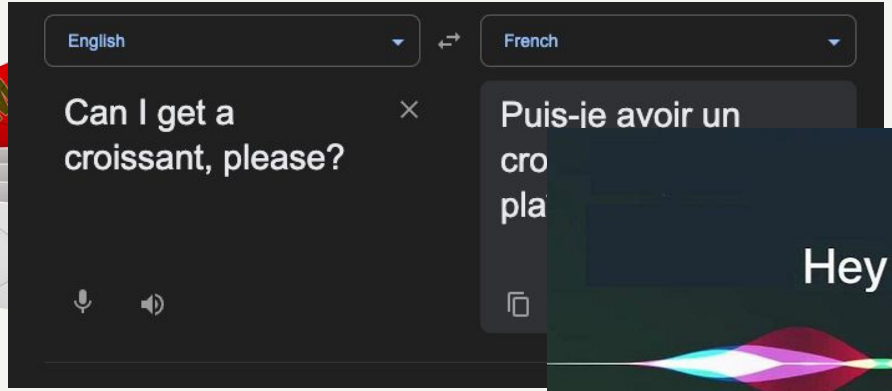
- Linguistics

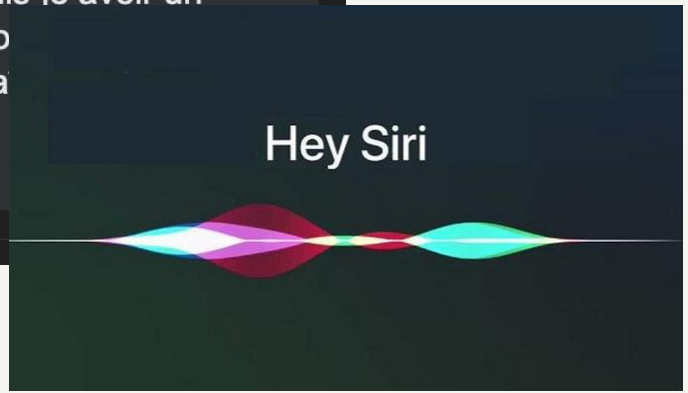
NLP is everywhere around us!

NLP is everywhere around us!

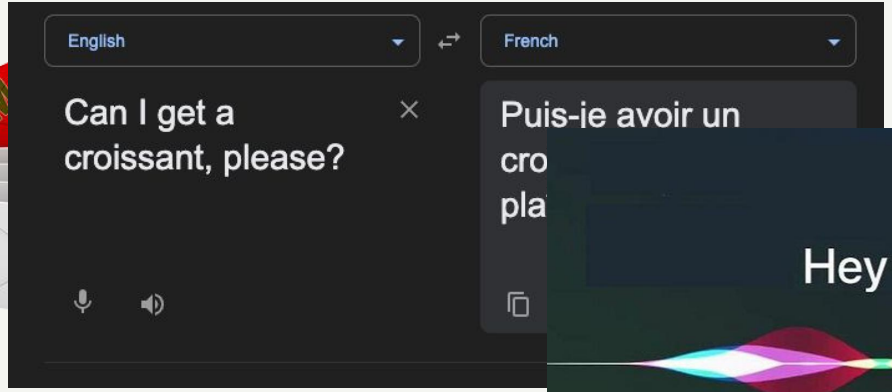


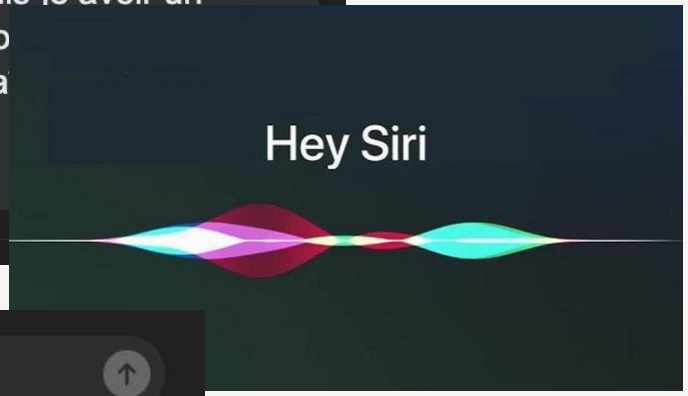
NLP is everywhere around us!

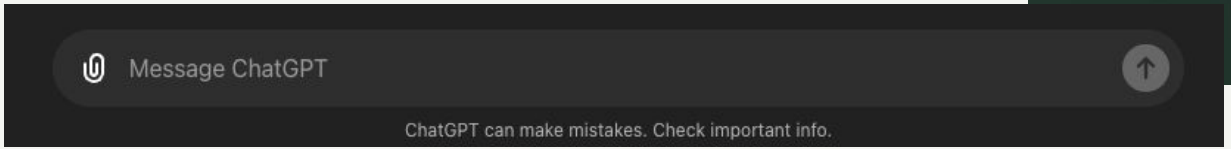




NLP is everywhere around us!

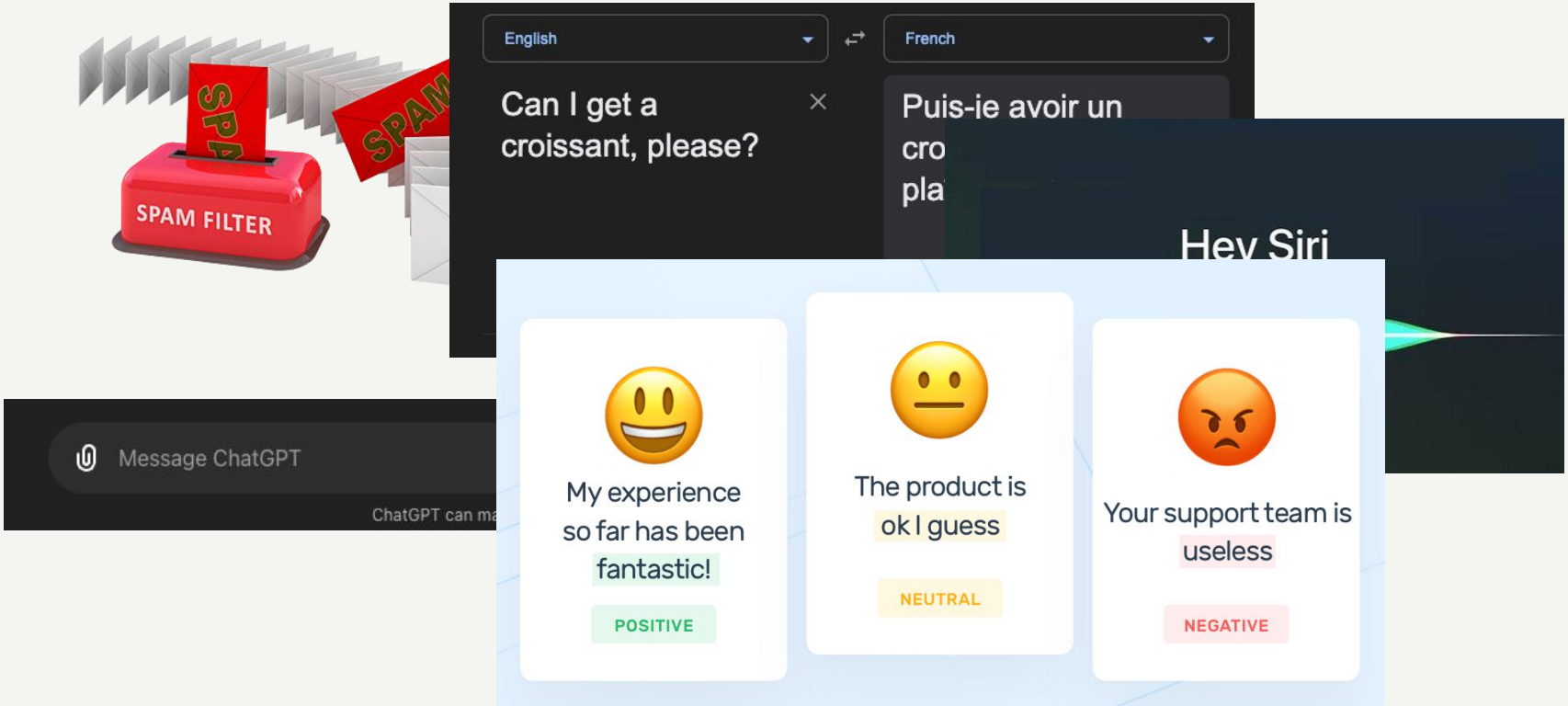






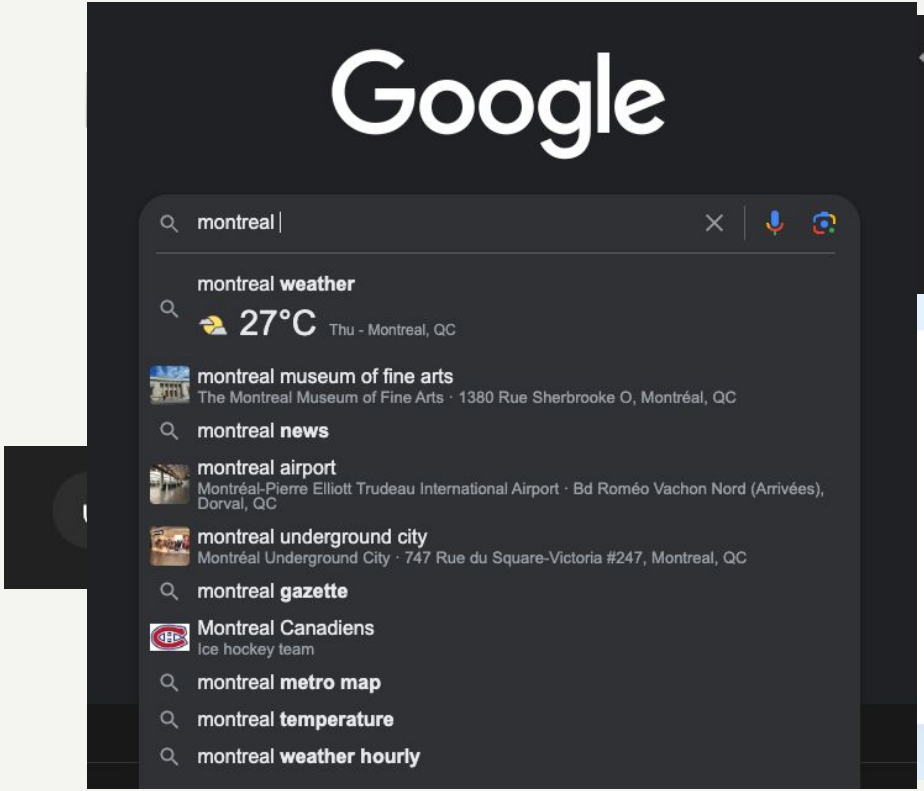
Mila

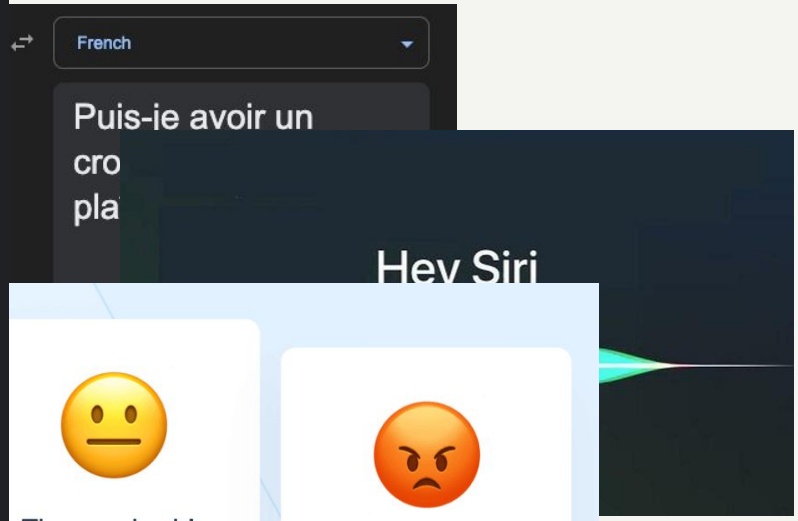
NLP is everywhere around us!

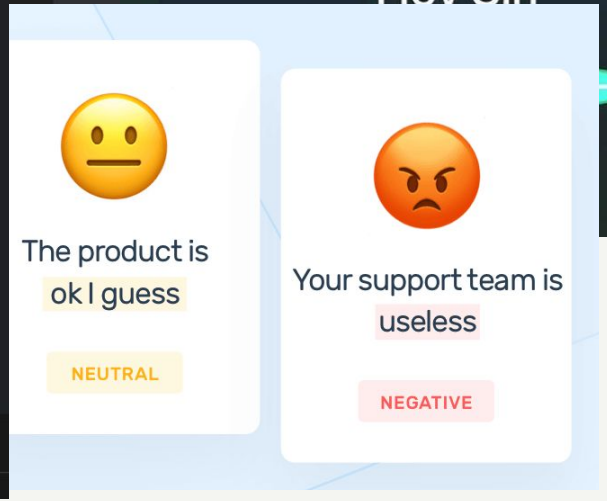


Mila

NLP is everywhere around us!







Mila

NLP is everywhere around us!

Mila

Why ‘natural language’? What other kind of language is there?

Why ‘natural language’? What other kind of language is there?

Natural Language

Source: https://www.thoughtco.com/ambiguity-language-1692388

Why ‘natural language’? What other kind of language is there?

Natural Language

Source: https://www.thoughtco.com/ambiguity-language-1692388

Computer Language

```
class Coder(BaseHuman):

  def __init__(self):
    coffee.strength++
    env.update()
    env.theme = DARK

  def day(self):
    self.eat(1*hrs)
    self.code(12*hrs)
    self.eat(1*hrs)
    self.debug(4*hrs)
    time.sleep(6*hrs)
```

Why ‘natural language’? What other kind of language is there?

Natural Language

- Used for everyday communication between people

Computer Language

- Used for instructing computers to perform specific tasks

Why ‘natural language’? What other kind of language is there?

Natural Language

- Used for everyday communication between people
- Developed naturally

Computer Language

- Used for instructing computers to perform specific tasks
- Systematically designed

Why ‘natural language’? What other kind of language is there?

Natural Language

- Used for everyday communication between people
- Developed naturally
- Complex and ambiguous

Computer Language

- Used for instructing computers to perform specific tasks
- Systematically designed
- Precise and unambiguous

Why ‘natural language’? What other kind of language is there?

Natural Language

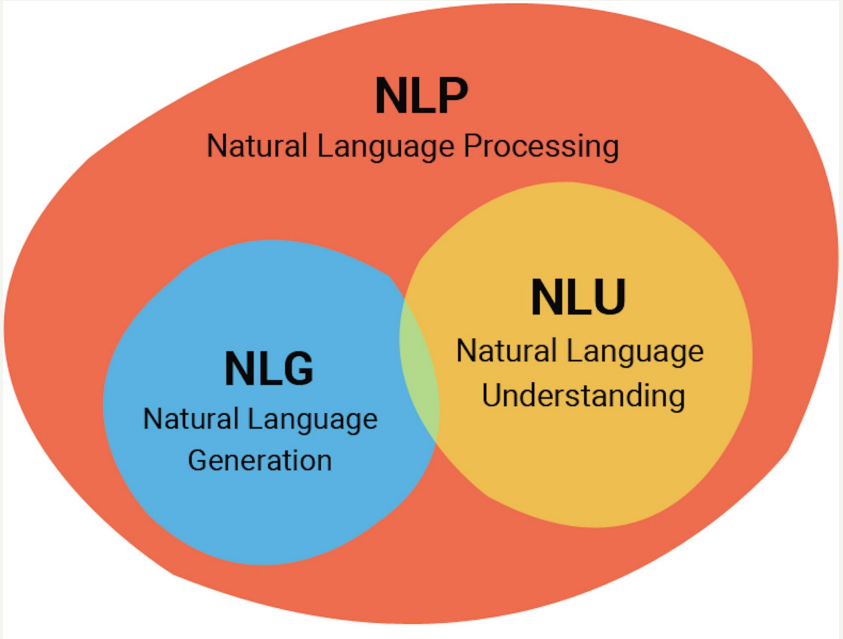
- Used for everyday communication between people
- Developed naturally
- Complex and ambiguous
- Highly nuanced and flexible

Computer Language

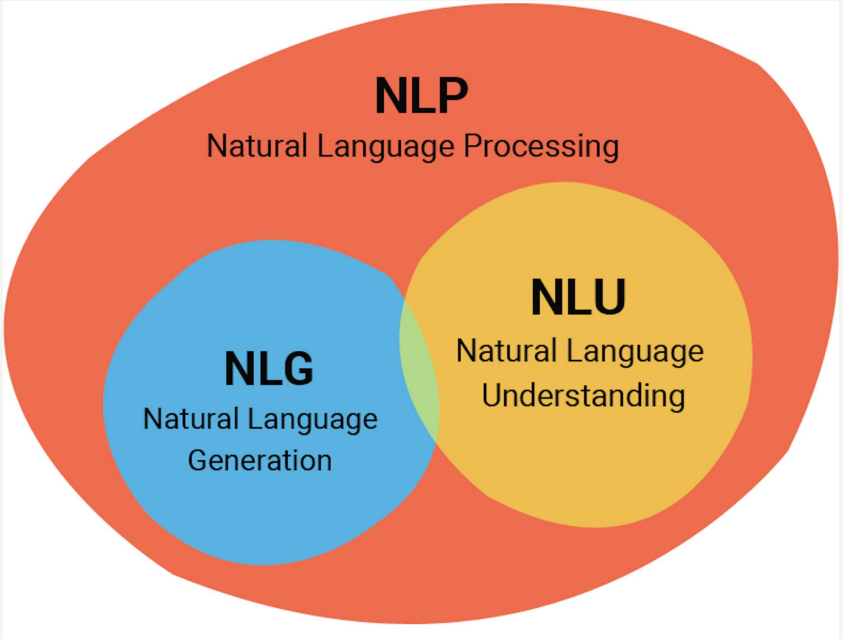
- Used for instructing computers to perform specific tasks
- Systematically designed
- Precise and unambiguous
- Limited in functionality and expressiveness

Terminology

NLP, NLU and NLG



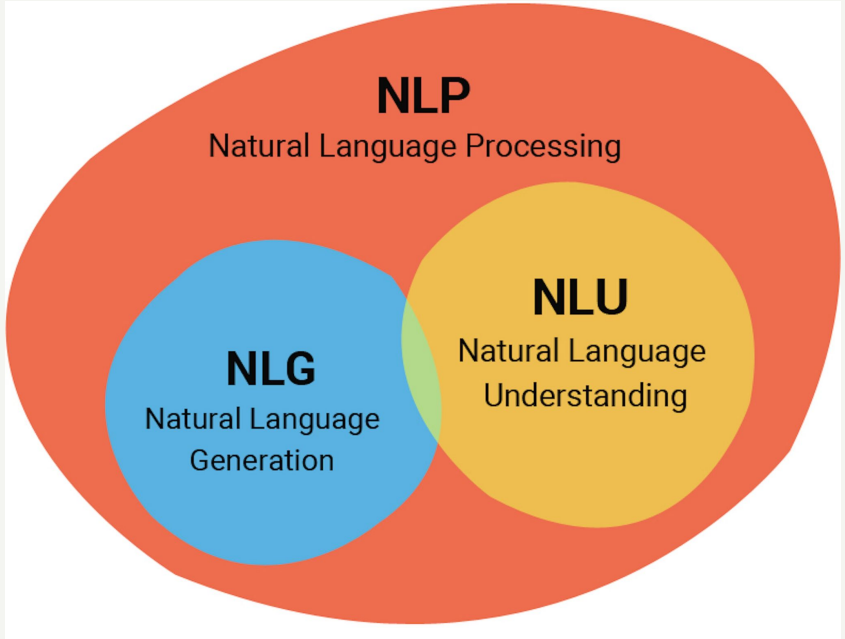
NLP, NLU and NLG



It enables computers to understand and interpret human language.

NLP, NLU and NLG

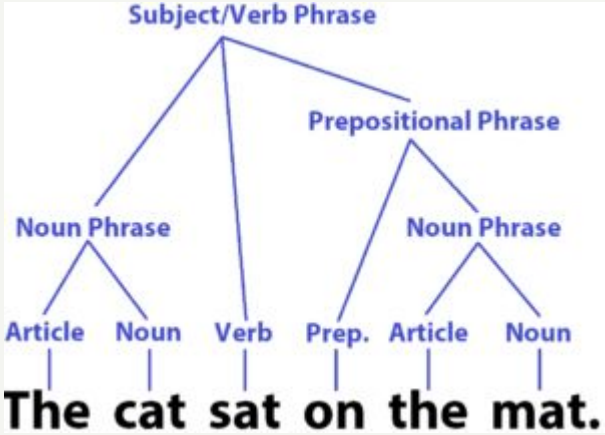
It enables computers to respond, manipulate and generate human language.

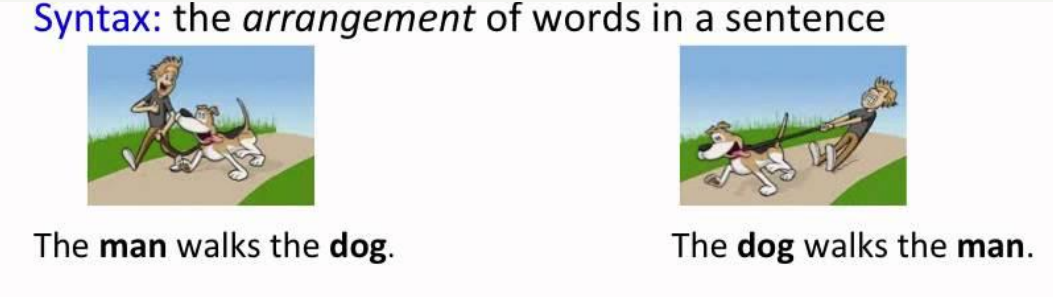


It enables computers to understand and interpret human language.

Syntax

sentence structure and grammar rules

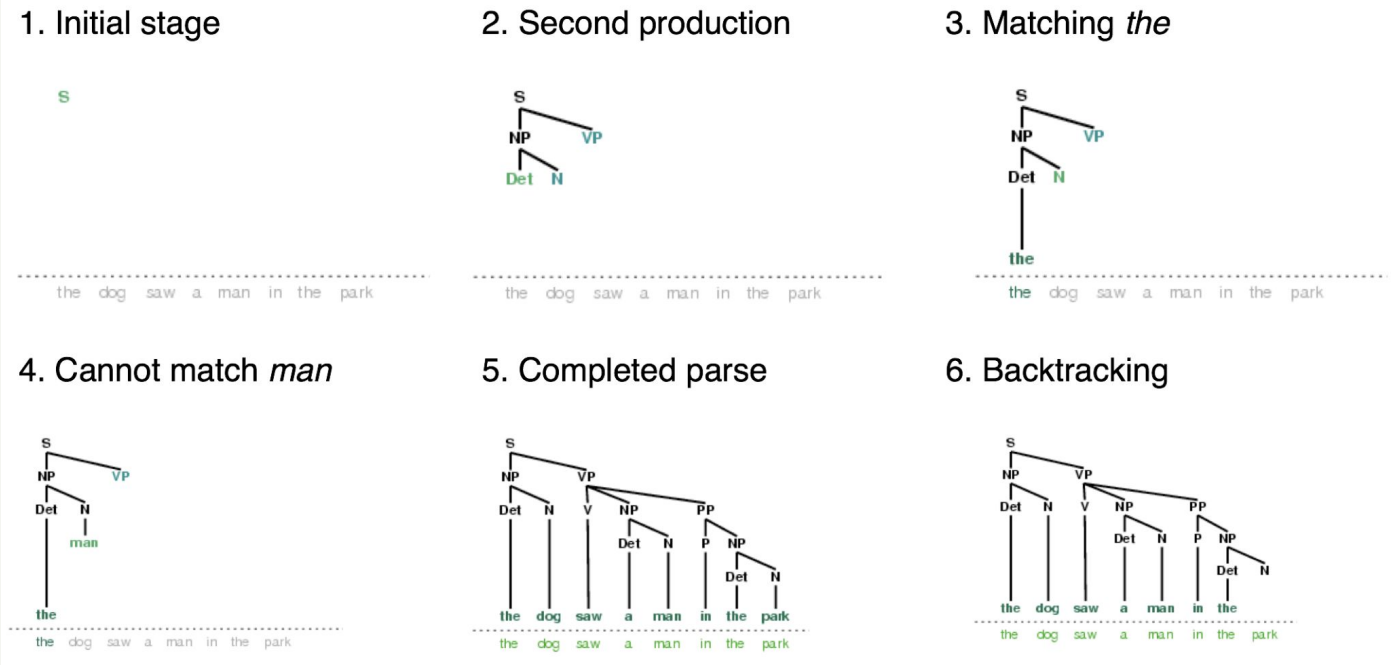




Source: https://www.youtube.com/watch?v=l3mbNkIEcYM

Parsing

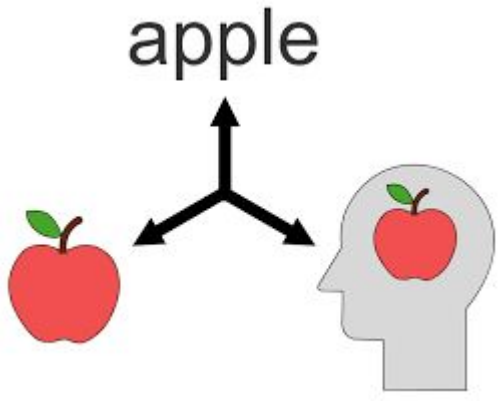
extracting syntax from a sentence



Source: https://www.nltk.org/book/ch08.html

Semantics

meaning of a word



Word	Semantic
pen	a writing tool
pen	a livestock's enclosure
pen	a portable enclosure for a baby
pen	a correctional institution
pen	a female swan

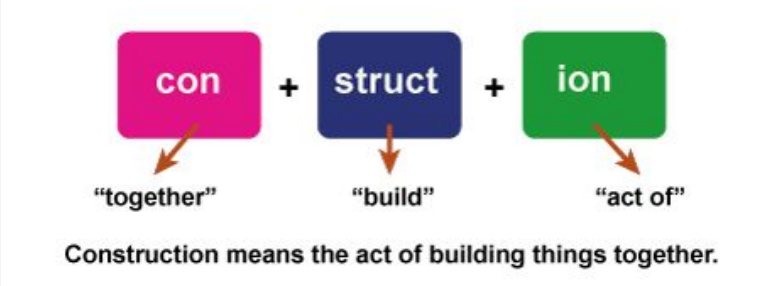
Pragmatics

meaning of a word in context of the sentence

It’s hot in here, can you crack a window?

Morphology

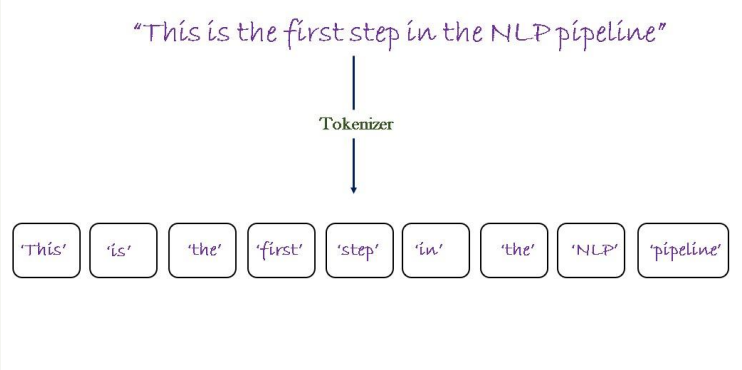
the study of how words are formed



Mila

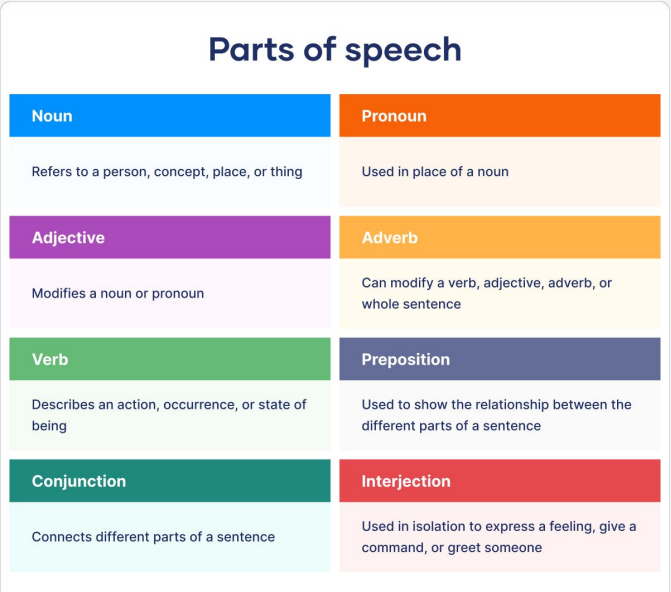
Tokenization

splitting text into smaller units (words, phrases, roots, etc.)



Parts-of-Speech (PoS)

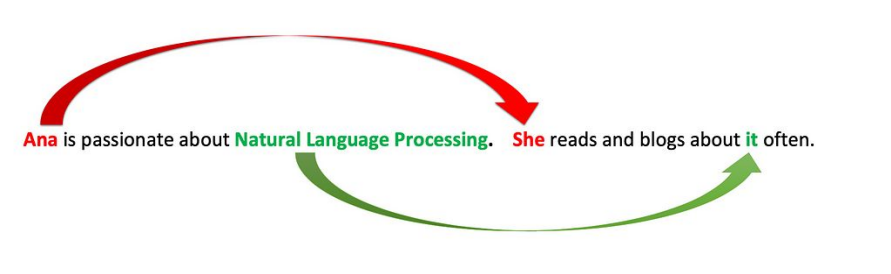
grammatical categories

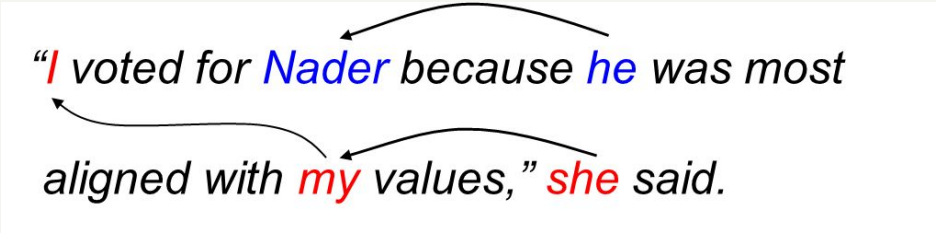


Source: https://www.scribbr.com/category/parts-of-speech/

Coreference

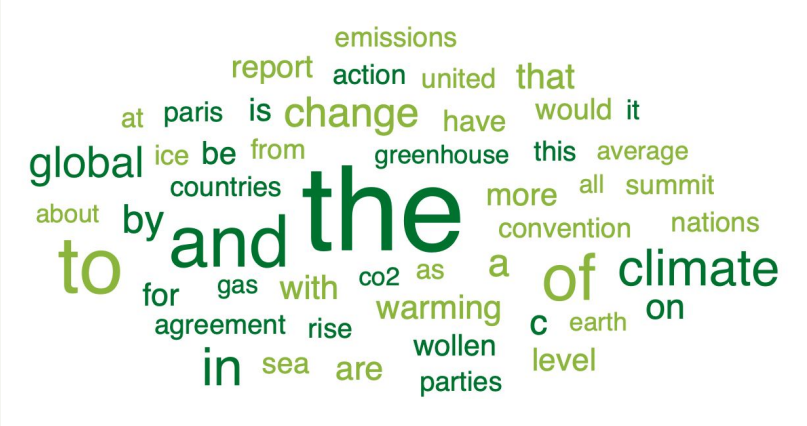
relationship between words which refer to the same entity

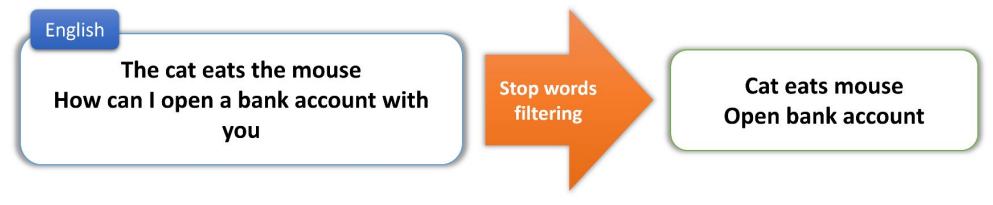




Stop Words

common words (like a, the, is, etc.) that are usually filtered out

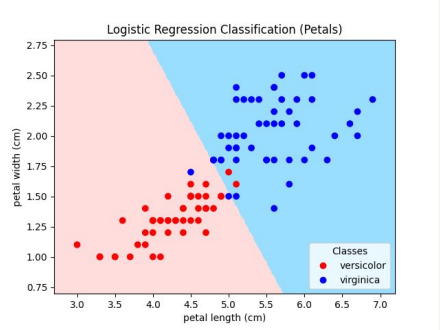




Mila

Applications of NLP

Applications of NLP: Text Classification



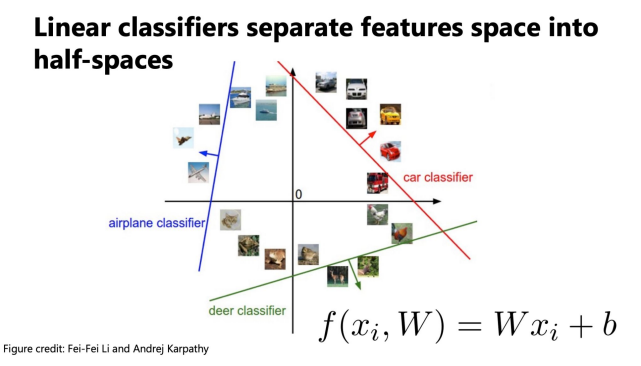
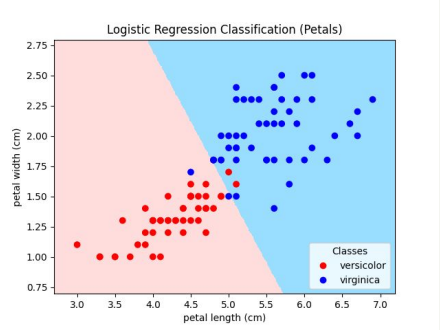
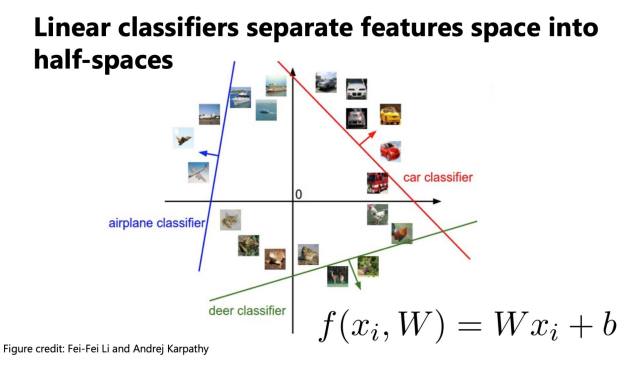


Figure credit: Fei-Fei Li and Andrej Karpathy

Remember classification with tabular data.
Classification with images.

Applications of NLP: Text Classification

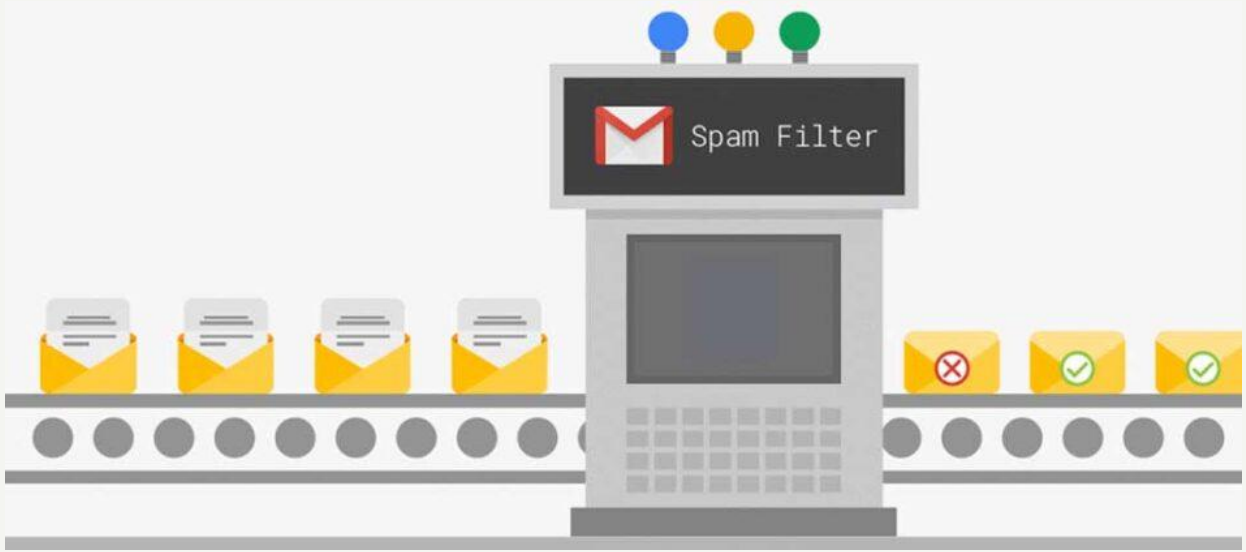




Remember classification with tabular data.
Classification with images.
Now we’ll do classification with natural language!

Applications of NLP: Text Classification

Filtering



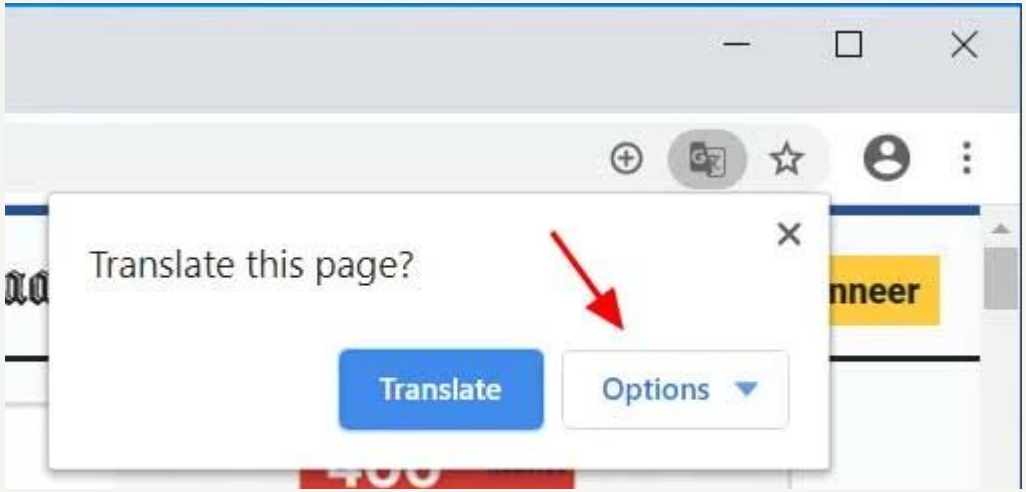
Applications of NLP: Text Classification

Sentiment Analysis and Topic Modeling

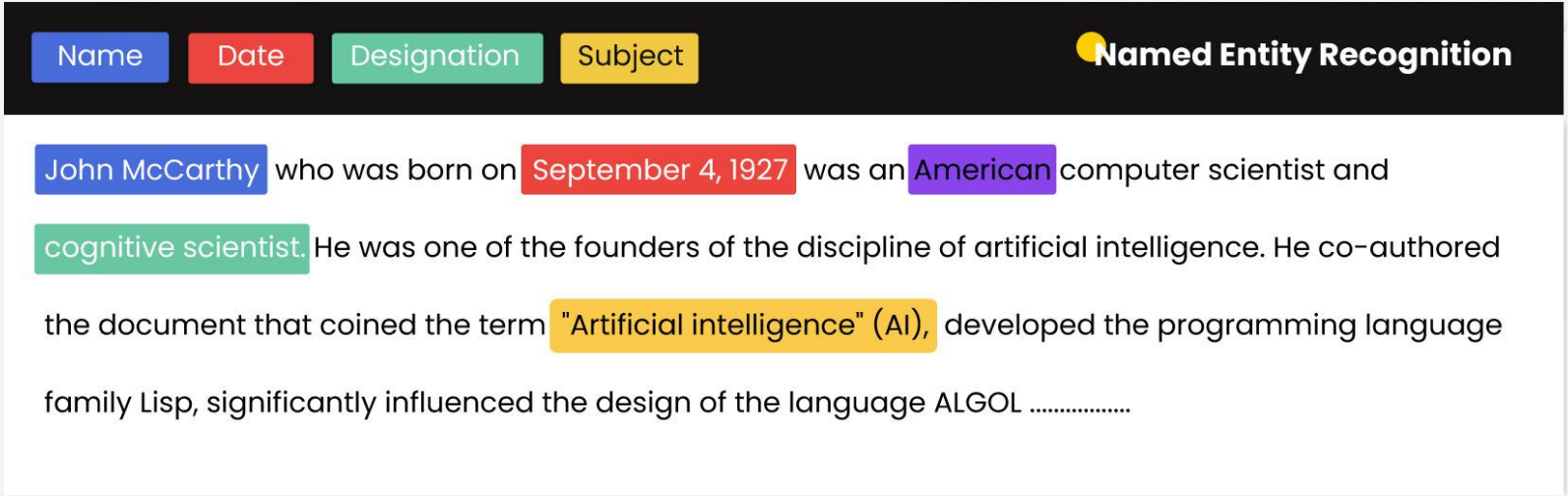
Mila

Applications of NLP: Text Classification

Language Identification



Applications of NLP: Named Entity Recognition



Applications of NLP: Machine Translation

Source: https://devnagri.com/machine-translation-important-things-to-know/

Applications of NLP: Vision Language Models

ChatGPT

What is going on in this picture?

This picture shows a meerkat standing on a digital scale, seemingly being weighed. The meerkat appears to be holding onto a clipboard held by a person who is likely a zookeeper or a veterinarian. The person is seated and holding a pen, possibly recording the meerkat's weight and other health-related information. The scene likely takes place in a zoo or an animal sanctuary during a routine health checkup for the meerkat. The meerkat's expression appears curious or playful, adding a humorous element to the image.

Applications of NLP: And so many others

1. Machine Translation
2. Speech Recognition
3. Text-to-Speech
4. Sentiment Analysis
5. Chatbots
6. Virtual Assistants
7. Named Entity Recognition (NER)
8. Part-of-Speech Tagging
9. Text Summarization
10. Question Answering
11. Document Classification
12. Information Retrieval
13. Spell Checking
14. Grammar Checking
15. Language Modeling
16. Topic Modeling
17. Text Generation
18. Emotion Detection
19. Plagiarism Detection
20. Keyword Extraction
21. Text Clustering
22. Language Detection
23. Autocomplete
24. Content Recommendation
25. Opinion Mining
26. Social Media Monitoring
27. Fake News Detection
28. Transliteration
29. Voice Assistants
30. Customer Support Automation
31. Market Intelligence
32. Fraud Detection
33. Medical Text Analysis
34. Legal Document Analysis
35. Automated Theorem Proving
36. Dialogue Systems
37. Semantic Search
38. Knowledge Graph Construction
39. Paraphrase Detection
40. Text Simplification
41. Email Filtering
42. Advertising and Marketing Analysis
43. Product Review Analysis
44. Machine Reading Comprehension
45. Optical Character Recognition (OCR)

.....

Challenges in NLP

Challenges of NLP: Phrasing Ambiguity

Source: https://blueskiesconsulting.com/how-well-do-you-handle-ambiguity-on-a-project/

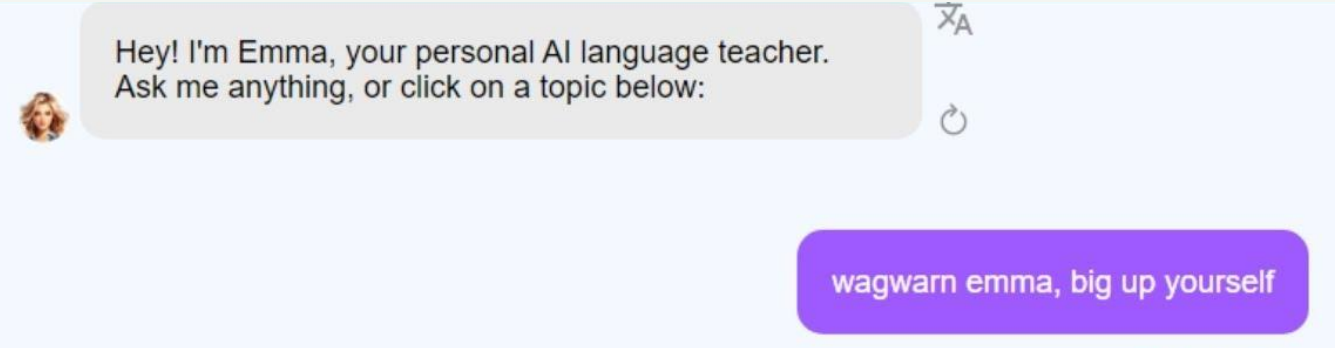
Challenges of NLP: Words with Multiple Meanings

Mila

Challenges of NLP: ~~Mispellings~~

Misspellings

Challenges of NLP: New Vocabulary

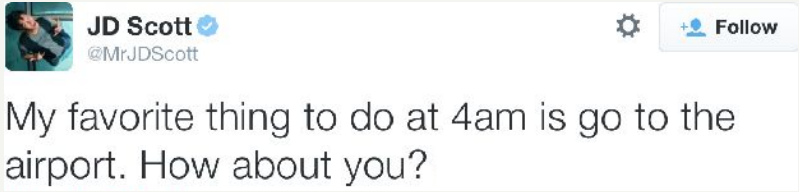


Mila

Challenges of NLP: Specialized Terminology

A 12-year old girl with known hyperagglutinability, presented to the emergency department with a 2-week history of headeaches and facial weakness. Neurologic examination indicated sensorineural hearing loss on the right side with Weber's test lateralizing to the left, and the Rinne's test demonstrating bone conduction greater than air conduction on the right. Magnetic resonance imaging of the head revealed severe structural defects of the right petrous temporal bone. No indication of cerebral infarction.

Challenges of NLP: Tone of Voice



Challenges of NLP: Understanding Context

It’s raining cats and dogs!

Source:
https://medium.com/@InsightfulScribbler/the-curious-history-of-raining-cats-and-dogs-and-interesting-rainy-weather-idioms-from-other-33709f6b7884

Challenges of NLP: Code Switching

Source: https://www.theinformedslp.com/review/a-little-bit-of-this-un-poquito-of-that

Break

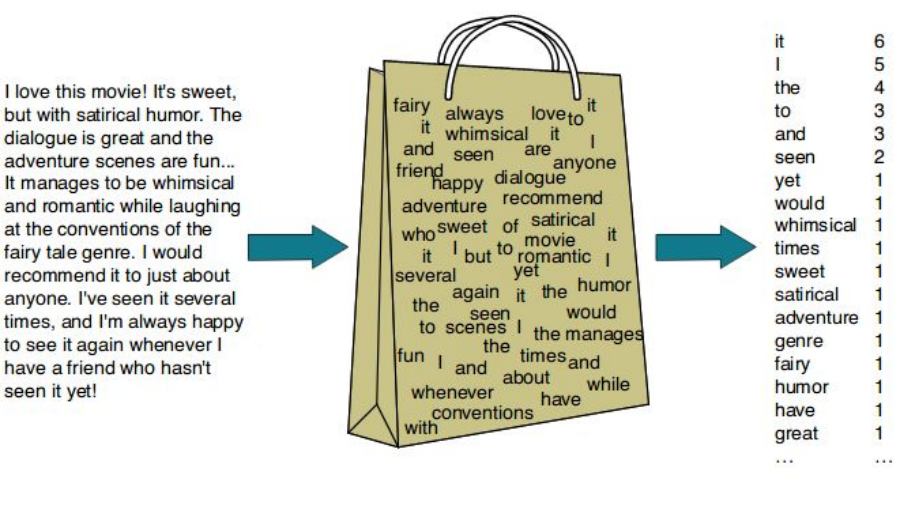
Modeling Language

Bag of Words

Order of the words doesn’t matter, only their occurrence matters.

Bag of Words

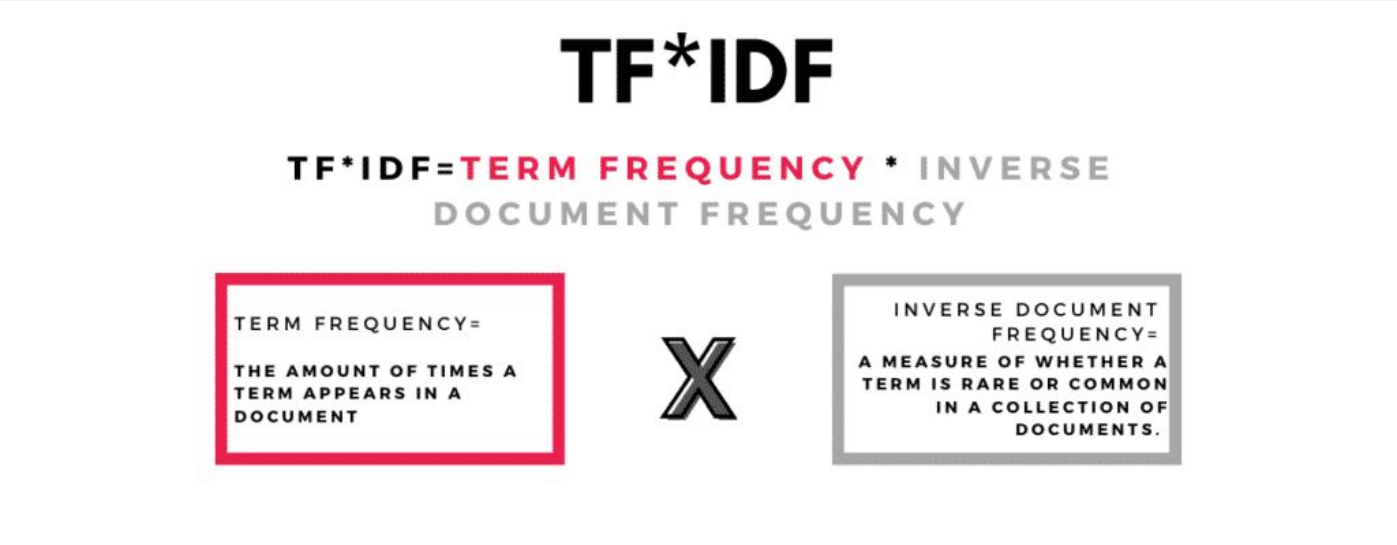
Order of the words doesn’t matter, only their occurrence matters.



Source: https://koushik1102.medium.com/nlp-bag-of-words-and-tf-idf-explained-fd1f49dce7c4

Bag of Words: TF-IDF

Order of the words doesn't matter, only their occurrence matters.



Bag of Words

Order of the words doesn’t matter, only their occurrence matters.

- Simple, efficient, and a decent baseline.
- **Ignores context and semantics!**

n-gram Models

Order of the **n-grams** doesn't matter, only their occurrence matters.

n-gram Models

Order of the **n-grams** doesn't matter, only their occurrence matters.

The cat sat on the mat.

Bag of Words

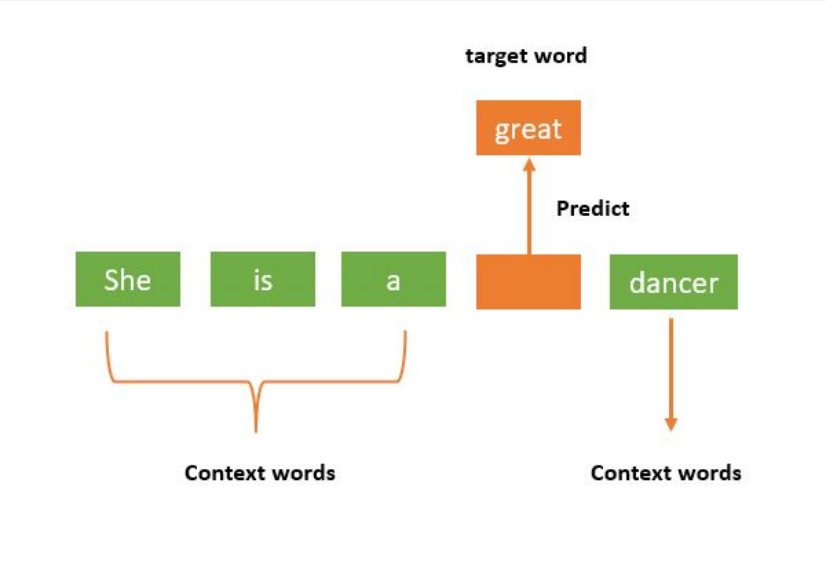
Bag of 2-grams

Continuous Bag of Words

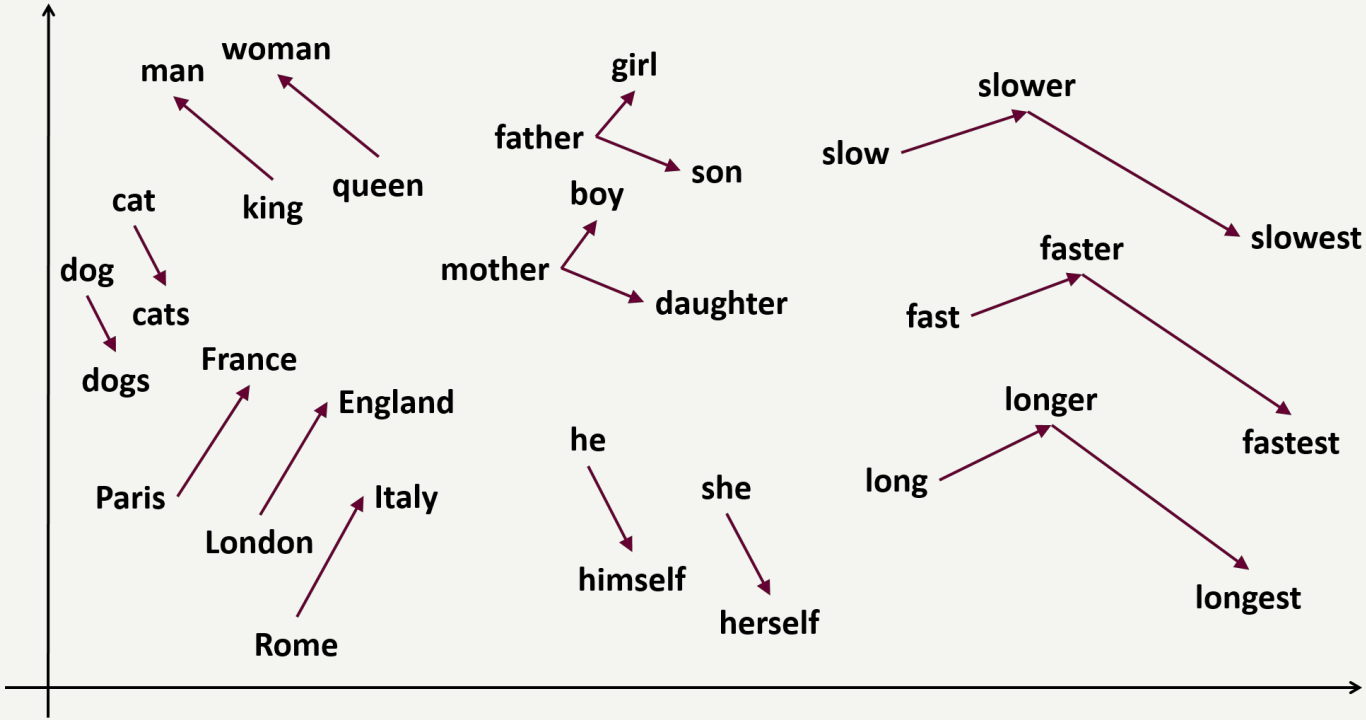
“You shall know a word by the company it keeps” - J.R. Firth

Continuous Bag of Words

"You shall know a word by the company it keeps" - J.R. Firth



Continuous Bag of Words: Word2Vec



Mila

Continuous Bag of Words: Word2Vec

https://projector.tensorflow.org/

Masked Language Modeling

Sentence:

The keys to the cabinet [MASK] on the table.

Mask 1 Predictions:

70.3% were
10.1% lay

Sentence:

The [MASK] to the cabinet were on the table.

Mask 1 Predictions:

89.7% keys
1.7% contents

Sentence:

The [MASK] to the cabinet [MASK] on the table.

Mask 1 Predictions:

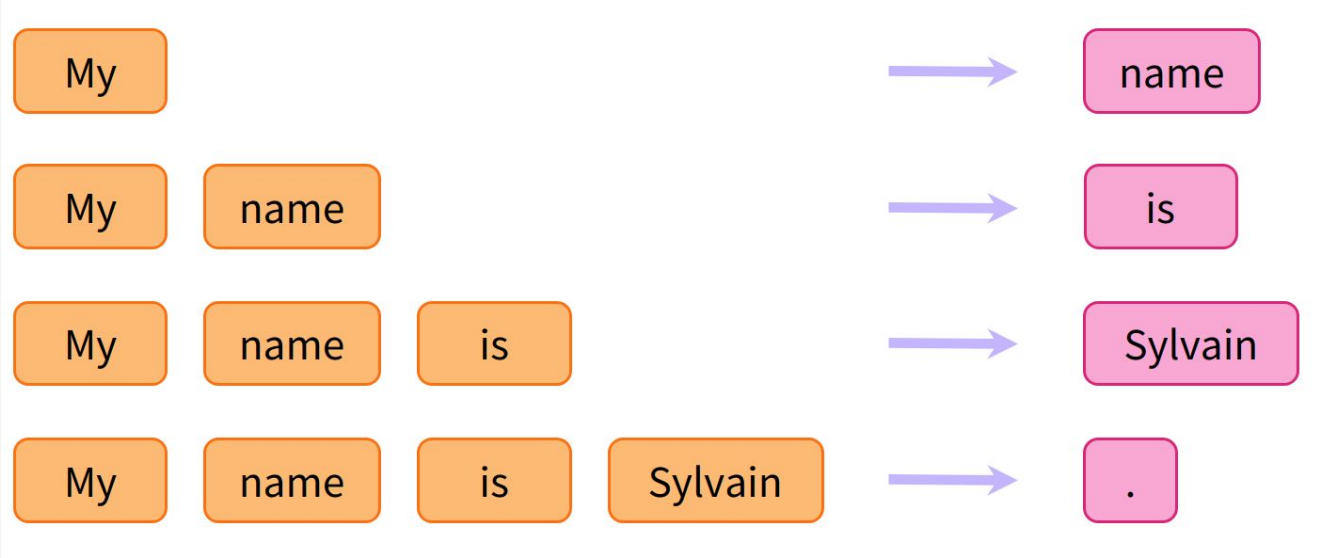
70.8% keys
18.2% key

Mask 2 Predictions:

36.6% was
9.0% were

Causal Language Modeling

Predicting the next word based on previous words.



Causal Language Modeling

NLP Pre-Deep Learning

Text Preprocessing: Cleaning up the Language

- **Lower Casing:** lOOK at that DUck! → look at that duck!

Text Preprocessing: Cleaning up the Language

- **Lower Casing:** lOOK at that DUck! → look at that duck!
- **Removing punctuations, stop words, special characters, etc.:** Holy sh!t, look at that duck!!! → look duck

Text Preprocessing: Cleaning up the Language

- **Lower Casing:** lOOK at that DUck! → look at that duck!
- **Removing punctuations, stop words, special characters, etc.:** Holy sh!t, look at that duck!!! → look duck
- **Normalization and Spell Correction:** U.K., UK, U K → United Kingdom

Text Preprocessing: Cleaning up the Language

- **Lower Casing:** lOOK at that DUck! → look at that duck!
- **Removing punctuations, stop words, special characters, etc.:** Holy sh!t, look at that duck!!! → look duck
- **Normalization and Spell Correction:** U.K., UK, U K → United Kingdom
- **Stemming and Lemmatization:** running → run
  fast, faster, fastest → fast

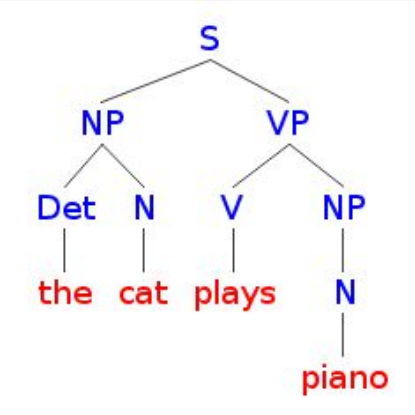
Text Preprocessing: Cleaning up the Language

- **Lower Casing:** lOOK at that DUck! → look at that duck!
- **Removing punctuations, stop words, special characters, etc.:** Holy sh!t, look at that duck!!! → look duck
- **Normalization and Spell Correction:** U.K., UK, U K → United Kingdom
- **Stemming and Lemmatization:** running → run
  fast, faster, fastest → fast

Just a lot of cleaning!

Other Parts of the Pipeline

- **Feature Extraction:** Bag of Words, TF-IDF, n-grams, etc.
- **Statistical Models:** HMMs, CRFs, Naive Bayes
- **Rule-based Parsing:** Syntax trees



Embeddings

Why Embeddings?

Why Embeddings?

Blood pressure = w*Dosage + b ⇒ **134 = 0.7*20 + 120**

Makes sense

Why Embeddings?

Blood pressure = w*Dosage + b ⇒ **134 = 0.7*20 + 120**

Makes sense

```
[112, 111,  98,  79,  97, 130, 124, 122, 127,  72]
[142, 124, 103, 104,  91,  92, 110, 100, 114,  65]
[137, 137, 119, 100,  98,  85,  98,  86,  94,  55]
[147, 142, 145, 129, 113,  99,  86,  81,  87,  62]
[143, 140, 141, 139, 137, 135, 153,  98,  87,  55]
[147, 151, 150, 148, 115, 163, 241, 170, 111,  82]
[152, 153, 152, 141,  60, 129, 202, 150, 127, 146]
[134, 142, 152,  99,  52,  90, 128, 107, 134, 148]
[136, 135, 131,  56,  74,  94, 119, 133, 144, 143]
[133, 138, 105,  50,  79,  87,  93, 137, 146, 145]
[131, 136,  90,  64,  80,  89,  80, 130, 135, 137]
[127, 125,  67,  80,  71,  85,  92, 134, 137, 131]
[118, 119,  48,  76,  73,  60,  88, 134, 133, 136]
[119, 114,  52, 100,  60,  10,  42, 101, 123, 132]
[114, 106,  81, 113,  22,  14,  59, 120, 131, 126]
[107, 109,  92,  65,  20,  64, 121, 125, 128, 134]
[107, 110,  35,  37,  75, 123, 136, 127, 124, 130]
[104, 121,  94, 111, 124, 124, 129, 130, 118, 124]
[102, 119, 127, 125, 122, 128, 129, 125, 126, 114]
[114, 117, 116, 115, 108, 116, 124, 120, 131, 128]
[113, 111, 108, 101, 104, 112, 112, 120, 118, 114]
[101, 105, 101,  91,  98, 107, 103, 109, 108, 109]
[ 98,  94,  96,  94,  93,  99,  90, 100, 110, 115]
[103,  99,  92,  91,  96,  98,  87,  89,  91, 103]
[102,  95,  90,  90,  92,  99,  91,  97,  95,  92]
```

"Applying a filter"

1	0	-1
2	0	-2
1	0	-1

filter

*

1	3	4
2	1	1
2	5	2

input

=

1*1 + 0*3 + (-1)*4

+ 2*2 + 0 *1 + (-2) *1 = -1

+ 1*2 + 0*5 + (-1)*2

Makes sense

Why Embeddings?

Blood pressure = w*Dosage + b ⇒ **134 = 0.7*20 + 120**

Makes sense

```
[112, 111,  98,  79,  97, 130, 124, 122, 127,  72]
[142, 124, 103, 104,  91,  92, 110, 100, 114,  65]
[137, 137, 119, 100,  98,  85,  98,  86,  94,  55]
[147, 142, 145, 129, 113,  99,  86,  81,  87,  62]
[143, 140, 141, 139, 137, 135, 153,  98,  87,  55]
[147, 151, 150, 148, 115, 163, 241, 170, 111,  82]
[152, 153, 152, 141,  60, 129, 202, 150, 127, 146]
[134, 142, 152,  99,  52,  90, 128, 107, 134, 148]
[136, 135, 131,  56,  74,  94, 119, 133, 144, 143]
[133, 138, 105,  50,  79,  87,  93, 137, 146, 145]
[131, 136,  90,  64,  80,  89,  80, 130, 135, 137]
[127, 125,  67,  80,  71,  85,  92, 134, 137, 131]
[118, 119,  48,  76,  73,  60,  88, 134, 133, 136]
[119, 114,  52, 100,  60,  10,  42, 101, 123, 132]
[114, 106,  81, 113,  22,  14,  59, 120, 131, 126]
[107, 109,  92,  65,  20,  64, 121, 125, 128, 134]
[107, 110,  35,  37,  75, 123, 136, 127, 124, 130]
[104, 121,  94, 111, 124, 124, 129, 130, 118, 124]
[102, 119, 127, 125, 122, 128, 129, 125, 126, 114]
[114, 117, 116, 115, 108, 116, 124, 120, 131, 128]
[113, 111, 108, 101, 104, 112, 112, 120, 118, 114]
[101, 105, 101,  91,  98, 107, 103, 109, 108, 109]
[ 98,  94,  96,  94,  93,  99,  90, 100, 110, 115]
[103,  99,  92,  91,  96,  98,  87,  89,  91, 103]
[102,  95,  90,  90,  92,  99,  91,  97,  95,  92]
```

"Applying a filter"

1	0	-1
2	0	-2
1	0	-1

filter

*

1	3	4
2	1	1
2	5	2

input

=

1*1 + 0*3 + (-1)*4

+ 2*2 + 0 *1 + (-2) *1 = -1

+ 1*2 + 0*5 + (-1)*2

Makes sense

The cat sat on the mat ⇒ The*0.7 + cat*1.3 + ...

????

Mila

Why Embeddings?

Blood pressure = w*Dosage + b ⇒ **134 = 0.7*20 + 120**

Makes sense

```
[112, 111,  98,  79,  97, 130, 124, 122, 127,  72]
[142, 124, 103, 104,  91,  92, 110, 100, 114,  65]
[137, 137, 119, 100,  98,  85,  98,  86,  94,  55]
[147, 142, 145, 129, 113,  99,  86,  81,  87,  62]
[143, 140, 141, 139, 137, 135, 153,  98,  87,  55]
[147, 151, 150, 148, 115, 163, 241, 170, 111,  82]
[152, 153, 152, 141,  60, 129, 202, 150, 127, 146]
[134, 142, 152,  99,  52,  90, 128, 107, 134, 148]
[136, 135, 131,  56,  74,  94, 119, 133, 144, 143]
[133, 138, 105,  50,  79,  87,  93, 137, 146, 145]
[131, 136,  90,  64,  80,  89,  80, 130, 135, 137]
[127, 125,  67,  80,  71,  85,  92, 134, 137, 131]
[118, 119,  48,  76,  73,  60,  88, 134, 133, 136]
[119, 114,  52, 100,  60,  10,  42, 101, 123, 132]
[114, 106,  81, 113,  22,  14,  59, 120, 131, 126]
[107, 109,  92,  65,  20,  64, 121, 125, 128, 134]
[107, 110,  35,  37,  75, 123, 136, 127, 124, 130]
[104, 121,  94, 111, 124, 124, 129, 130, 118, 124]
[102, 119, 127, 125, 122, 128, 129, 125, 126, 114]
[114, 117, 116, 115, 108, 116, 124, 120, 131, 128]
[113, 111, 108, 101, 104, 112, 112, 120, 118, 114]
[101, 105, 101,  91,  98, 107, 103, 109, 108, 109]
[ 98,  94,  96,  94,  93,  99,  90, 100, 110, 115]
[103,  99,  92,  91,  96,  98,  87,  89,  91, 103]
[102,  95,  90,  90,  92,  99,  91,  97,  95,  92]
```

"Applying a filter"

1	0	-1
2	0	-2
1	0	-1

filter

*

1	3	4
2	1	1
2	5	2

input

=

1*1 + 0*3 + (-1)*4

+ 2*2 + 0 *1 + (-2) *1 = -1

+ 1*2 + 0*5 + (-1)*2

Makes sense

We need a way to numerically represent language

The cat sat on the mat ⇒ The*0.7 + cat*1.3 + ...

????

Mila

Embeddings as Sequential Numbering

The	cat	sat	on	the	mat
1	2	3	4	1	5

Will this work?

Embeddings as Sequential Numbering

The cat sat on the mat
1 2 3 4 1 5

Will this work?

Are the words ‘the’ and ‘cat’ similar? 2-1 = 1. Yes
Are the words ‘the’ and ‘mat’ similar? 5-1 = 4. No

We have encoded *wrong* similarity information into these embeddings without even wanting to!

Embeddings as One Hot Encoding

$$\begin{pmatrix} the \\ cat \\ sat \\ on \\ the \\ mat \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Is this better?

Embeddings as One Hot Encoding

$$\begin{pmatrix} the \\ cat \\ sat \\ on \\ the \\ mat \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Is this better?

Better. Distance or ‘similarity’ between any 2 feature vectors is now the same! But we’re not done yet.

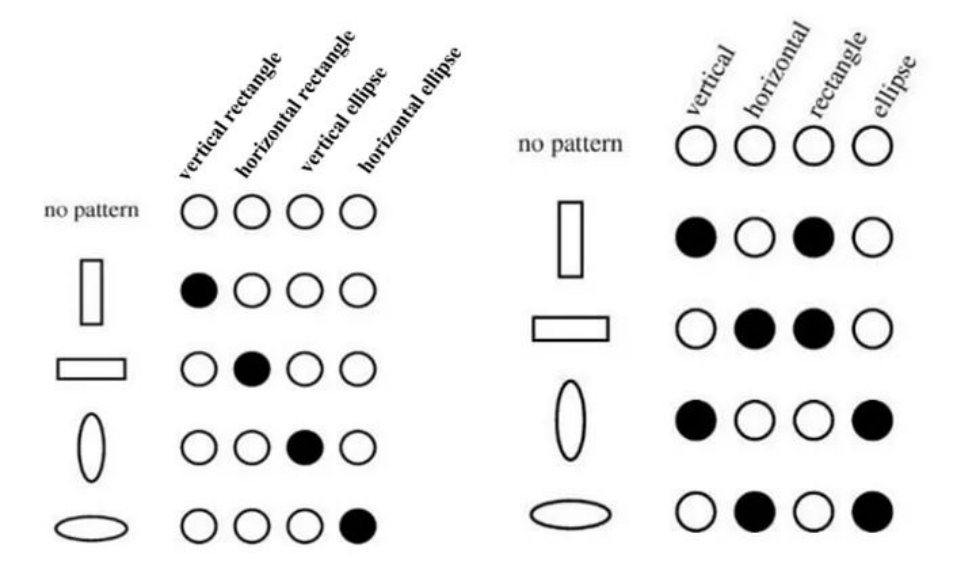
This representation does not have the problems of sequential numbering but it also **holds no similarity information** about the relationship between words.

Embeddings as Distributed Representation

Numerical representation with **correct** comparative value!

Embeddings as Distributed Representation

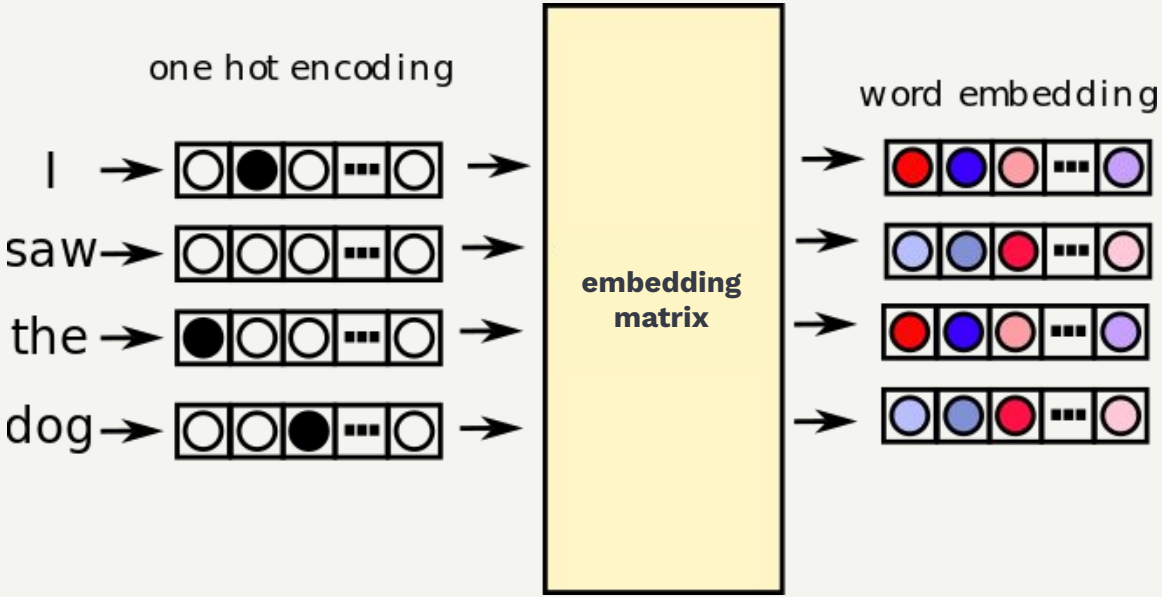
Numerical representation with **correct** comparative value!



One hot encoding

Distributed Representation

Embedding Matrix



Embedding Matrix

the	1	0	0	0	0
cat	0	1	0	0	0
sat	0	0	1	0	0
on	0	0	0	1	0
the	1	0	0	0	0
mat	0	0	0	0	1

One hot encoding

Embedding Matrix

the	1	0	0	0	0
cat	0	1	0	0	0
sat	0	0	1	0	0
on	0	0	0	1	0
the	1	0	0	0	0
mat	0	0	0	0	1

One hot encoding

$\times$

Embedding Size = 3

0.3	0.7	0.9
0.2	0.8	1.1
0.4	0.5	0.5
0.4	0.3	1.2
0.7	0.8	0.9

Embedding Matrix
(Learnable)

Embedding Matrix

the	1	0	0	0	0
cat	0	1	0	0	0
sat	0	0	1	0	0
on	0	0	0	1	0
the	1	0	0	0	0
mat	0	0	0	0	1

One hot encoding

×

Embedding Size = 3

0.3	0.7	0.9
0.2	0.8	1.1
0.4	0.5	0.5
0.4	0.3	1.2
0.7	0.8	0.9

Embedding Matrix (Learnable)

=

the			
cat			
sat			
on			
the			
mat			

Distributed Representation (Embeddings)

Embedding Matrix

the	1	0	0	0	0
cat	0	1	0	0	0
sat	0	0	1	0	0
on	0	0	0	1	0
the	1	0	0	0	0
mat	0	0	0	0	1

One hot encoding

×

Embedding Size = 3

0.3	0.7	0.9
0.2	0.8	1.1
0.4	0.5	0.5
0.4	0.3	1.2
0.7	0.8	0.9

Embedding Matrix (Learnable)

=

the			
cat			
sat			
on			
the			
mat			

Distributed Representation (Embeddings)

Embedding Matrix

the	1	0	0	0	0
cat	0	1	0	0	0
sat	0	0	1	0	0
on	0	0	0	1	0
the	1	0	0	0	0
mat	0	0	0	0	1

One hot encoding

×

Embedding Size = 3

0.3	0.7	0.9
0.2	0.8	1.1
0.4	0.5	0.5
0.4	0.3	1.2
0.7	0.8	0.9

Embedding Matrix (Learnable)

=

the	0.3		
cat			
sat			
on			
the			
mat			

Distributed Representation (Embeddings)

Embedding Matrix

the	1	0	0	0	0
cat	0	1	0	0	0
sat	0	0	1	0	0
on	0	0	0	1	0
the	1	0	0	0	0
mat	0	0	0	0	1

One hot encoding

×

Embedding Size = 3

0.3	0.7	0.9
0.2	0.8	1.1
0.4	0.5	0.5
0.4	0.3	1.2
0.7	0.8	0.9

Embedding Matrix (Learnable)

=

the	0.3		
cat			
sat			
on			
the			
mat			

Distributed Representation (Embeddings)

Embedding Matrix

the	1	0	0	0	0
cat	0	1	0	0	0
sat	0	0	1	0	0
on	0	0	0	1	0
the	1	0	0	0	0
mat	0	0	0	0	1

One hot encoding

×

Embedding Size = 3

0.3	0.7	0.9
0.2	0.8	1.1
0.4	0.5	0.5
0.4	0.3	1.2
0.7	0.8	0.9

Embedding Matrix (Learnable)

=

the	0.3		
cat	0.2		
sat			
on			
the			
mat			

Distributed Representation (Embeddings)

Embedding Matrix

the	1	0	0	0	0
cat	0	1	0	0	0
sat	0	0	1	0	0
on	0	0	0	1	0
the	1	0	0	0	0
mat	0	0	0	0	1

One hot encoding

×

Embedding Size = 3

0.3	0.7	0.9
0.2	0.8	1.1
0.4	0.5	0.5
0.4	0.3	1.2
0.7	0.8	0.9

Embedding Matrix (Learnable)

=

the	0.3		
cat	0.2		
sat	0.4		
on	0.4		
the	0.3		
mat	0.7		

Distributed Representation (Embeddings)

Embedding Matrix

the	1	0	0	0	0
cat	0	1	0	0	0
sat	0	0	1	0	0
on	0	0	0	1	0
the	1	0	0	0	0
mat	0	0	0	0	1

One hot encoding

×

Embedding Size = 3

0.3	0.7	0.9
0.2	0.8	1.1
0.4	0.5	0.5
0.4	0.3	1.2
0.7	0.8	0.9

Embedding Matrix (Learnable)

=

the	0.3		
cat	0.2		
sat	0.4		
on	0.4		
the	0.3		
mat	0.7		

Distributed Representation (Embeddings)

Embedding Matrix

the	1	0	0	0	0
cat	0	1	0	0	0
sat	0	0	1	0	0
on	0	0	0	1	0
the	1	0	0	0	0
mat	0	0	0	0	1

One hot encoding

×

Embedding Size = 3

0.3	0.7	0.9
0.2	0.8	1.1
0.4	0.5	0.5
0.4	0.3	1.2
0.7	0.8	0.9

Embedding Matrix (Learnable)

=

the	0.3	0.7	
cat	0.2		
sat	0.4		
on	0.4		
the	0.3		
mat	0.7		

Distributed Representation (Embeddings)

Embedding Matrix

the	1	0	0	0	0
cat	0	1	0	0	0
sat	0	0	1	0	0
on	0	0	0	1	0
the	1	0	0	0	0
mat	0	0	0	0	1

One hot encoding

×

Embedding Size = 3

0.3	0.7	0.9
0.2	0.8	1.1
0.4	0.5	0.5
0.4	0.3	1.2
0.7	0.8	0.9

Embedding Matrix (Learnable)

=

the	0.3	0.7	
cat	0.2		
sat	0.4		
on	0.4		
the	0.3		
mat	0.7		

Distributed Representation (Embeddings)

Embedding Matrix

the	1	0	0	0	0
cat	0	1	0	0	0
sat	0	0	1	0	0
on	0	0	0	1	0
the	1	0	0	0	0
mat	0	0	0	0	1

One hot encoding

×

Embedding Size = 3

0.3	0.7	0.9
0.2	0.8	1.1
0.4	0.5	0.5
0.4	0.3	1.2
0.7	0.8	0.9

Embedding Matrix (Learnable)

=

the	0.3	0.7	
cat	0.2	0.8	
sat	0.4		
on	0.4		
the	0.3		
mat	0.7		

Distributed Representation (Embeddings)

Embedding Matrix

the	1	0	0	0	0
cat	0	1	0	0	0
sat	0	0	1	0	0
on	0	0	0	1	0
the	1	0	0	0	0
mat	0	0	0	0	1

One hot encoding

×

Embedding Size = 3

0.3	0.7	0.9
0.2	0.8	1.1
0.4	0.5	0.5
0.4	0.3	1.2
0.7	0.8	0.9

Embedding Matrix (Learnable)

=

the	0.3	0.7	0.9
cat	0.2	0.8	1.1
sat	0.4	0.5	0.5
on	0.4	0.3	1.2
the	0.3	0.7	0.9
mat	0.7	0.8	0.9

Distributed Representation (Embeddings)

Embedding Matrix

the	1	0	0	0	0
cat	0	1	0	0	0
sat	0	0	1	0	0
on	0	0	0	1	0
the	1	0	0	0	0
mat	0	0	0	0	1

One hot encoding

×

Embedding Size = 3

0.3	0.7	0.9
0.2	0.8	1.1
0.4	0.5	0.5
0.4	0.3	1.2
0.7	0.8	0.9

Embedding Matrix (Learnable)

=

the	0.3	0.7	0.9
cat	0.2	0.8	1.1
sat	0.4	0.5	0.5
on	0.4	0.3	1.2
the	0.3	0.7	0.9
mat	0.7	0.8	0.9

Distributed Representation (Embeddings)

Embedding Matrix

the	1	0	0	0	0
cat	0	1	0	0	0
sat	0	0	1	0	0
on	0	0	0	1	0
the	1	0	0	0	0
mat	0	0	0	0	1

One hot encoding

×

Embedding Size = 3

0.3	0.7	0.9
0.2	0.8	1.1
0.4	0.5	0.5
0.4	0.3	1.2
0.7	0.8	0.9

Embedding Matrix (Learnable)

=

the	0.3	0.7	0.9
cat	0.2	0.8	1.1
sat	0.4	0.5	0.5
on	0.4	0.3	1.2
the	0.3	0.7	0.9
mat	0.7	0.8	0.9

Distributed Representation (Embeddings)

Embedding Matrix

the	1	0	0	0	0
cat	0	1	0	0	0
sat	0	0	1	0	0
on	0	0	0	1	0
the	1	0	0	0	0
mat	0	0	0	0	1

One hot encoding

×

Embedding Size = 3

0.3	0.7	0.9
0.2	0.8	1.1
0.4	0.5	0.5
0.4	0.3	1.2
0.7	0.8	0.9

Embedding Matrix (Learnable)

=

the	0.3	0.7	0.9
cat	0.2	0.8	1.1
sat	0.4	0.5	0.5
on	0.4	0.3	1.2
the	0.3	0.7	0.9
mat	0.7	0.8	0.9

Distributed Representation (Embeddings)

Embedding Matrix

The embedding matrix is simply a representation of different words in the distributed representation space!

the	1	0	0	0	0
cat	0	1	0	0	0
sat	0	0	1	0	0
on	0	0	0	1	0
the	1	0	0	0	0
mat	0	0	0	0	1

One hot encoding

×

Embedding Size = 3

the	0.3	0.7	0.9
cat	0.2	0.8	1.1
sat	0.4	0.5	0.5
on	0.4	0.3	1.2
mat	0.7	0.8	0.9

Embedding Matrix (Learnable)

=

the	0.3	0.7	0.9
cat	0.2	0.8	1.1
sat	0.4	0.5	0.5
on	0.4	0.3	1.2
the	0.3	0.7	0.9
mat	0.7	0.8	0.9

Distributed Representation (Embeddings)

Embedding Matrix

Once we have numerical representation of the language, we can use the learning methods we studied earlier.

And some special methods designed just for NLP!

Embedding Matrix

Once we have numerical representation of the language, we can use the learning methods we studied earlier.

And some special methods designed just for NLP!

In the next class