

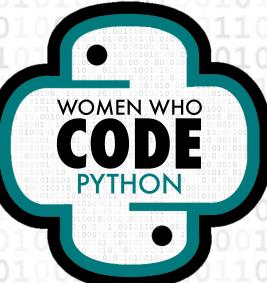
# Discover NLP with Python



Welcome to Session #4!

# Welcome everyone!

- You can find these slides on GitHub here:  
<https://github.com/WomenWhoCode/WWCodePython>
- Please make sure your chat is set to “All panelists and attendees”.
- Some housekeeping rules:
  - Everyone will be muted throughout the webinar, but there will be opportunities for participation!
  - Please share your thoughts on the chat and/or ask questions in the Q&A.
  - The entire team is here today. Please reach out to us with any technical questions!



# THANK YOU TO OUR LEADERS!



# OUR MISSION

Inspiring women to  
excel in technology  
careers.

WOMEN WHO  
**CODE**



# OUR VISION

A world where women are representative as technical executives, founders, VCs, board members and software engineers.



# OUR TARGET

Engineers with two or more years of experience looking for support and resources to strengthen their influence and levelup in their careers.



# CODE OF CONDUCT

**WWCode is an inclusive community**, dedicated to providing an empowering experience for everyone who participates in or supports our community, regardless of gender, gender identity and expression, sexual orientation, ability, physical appearance, body size, race, ethnicity, age, religion, socioeconomic status, caste, creed, political affiliation, or preferred programming language(s).

Our events are intended to inspire women to excel in technology careers, and anyone who is there for this purpose is welcome. We do not tolerate harassment of members in any form. Our **Code of Conduct** applies to all WWCode events and online communities.

Read the full version and access our incident report form at [womenwhocode.com/codeofconduct](http://womenwhocode.com/codeofconduct)



# 230,000

## Members

70 networks in 20 countries

Members in 97+ countries

10K+ events

\$1025 daily Conference tickets

\$2M Scholarships

Access to [jobs](#) + [resources](#)

Infinite connections



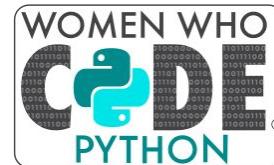
# OUR MOVEMENT

As the world changes, we can be a connecting force that creates a sense of belonging while the world is being asked to isolate.



# Neural Models for machine translation/conversion

*Session 4 : Intro to statistical machine translation  
and neural models for translation and conversion*



# AGENDA

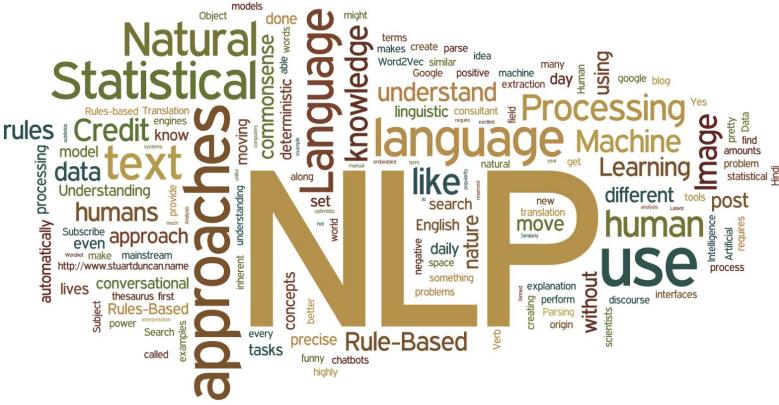
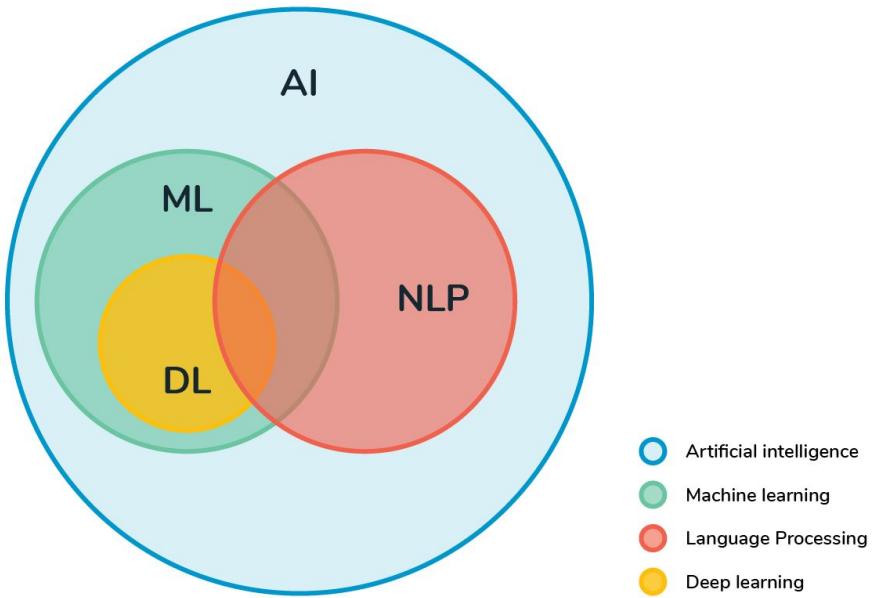
1. Recap - Week 1-3
  2. What is Machine Translation?
  3. Brief History of MT
  4. What is Statistical Machine Translation
  5. What is Neural Machine Translation
    - Sequence to Sequence Modeling
  6. Encoder Decoder Model
  7. Google Collab - Live Coding
-

1.

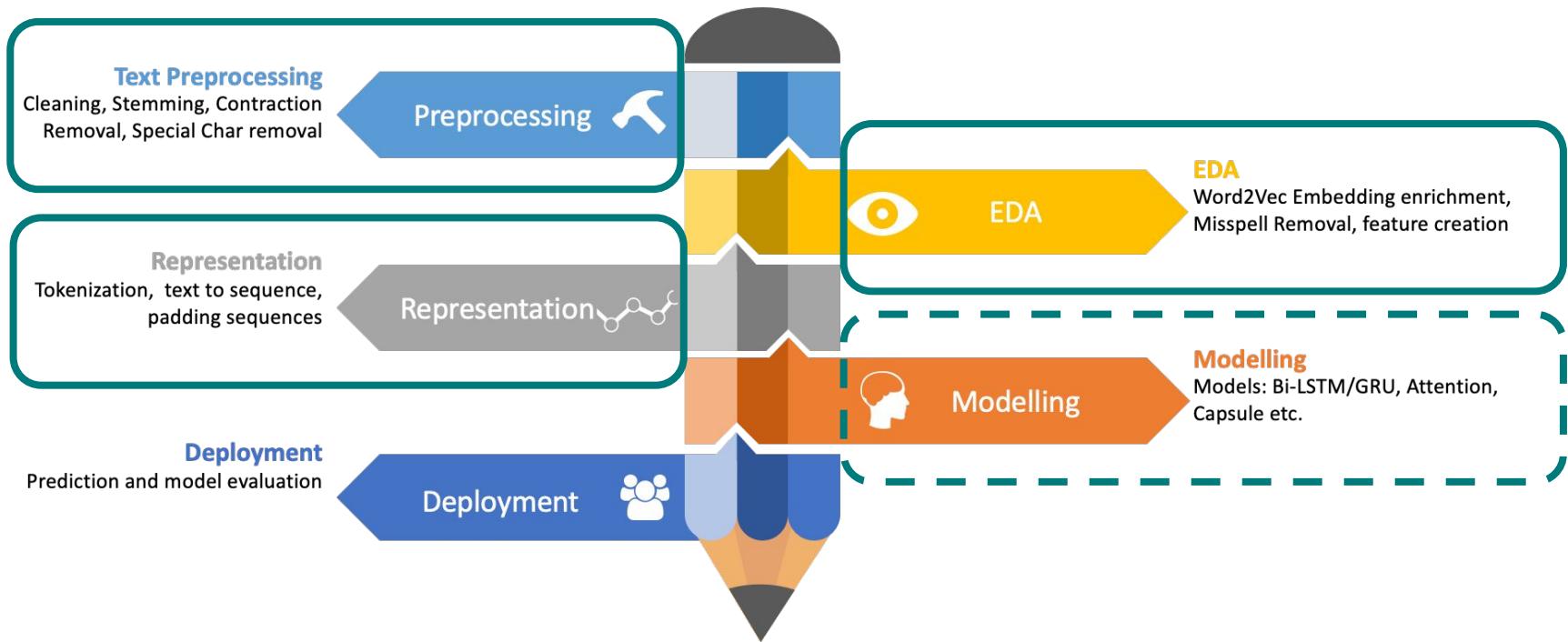
# Recap: Sessions #1-3



# Natural Language Processing



# The NLP Pipeline

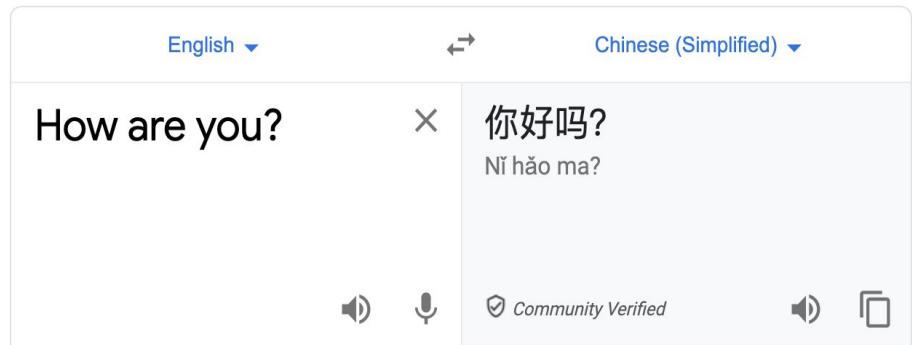


2.

# Machine Translation



# What is Machine Translation?



The screenshot shows a machine translation interface. At the top, it displays language pairs: English (dropdown) ↔ Chinese (Simplified) (dropdown). Below this, the English input "How are you?" is on the left, and its Chinese translation "你好吗?" (Nǐ hǎo ma?) is on the right. A small "X" icon is between the two inputs. At the bottom, there are four icons: a speaker icon, a microphone icon, a shield icon labeled "Community Verified", and a square icon.

- Sub-field of **computational linguistics** that investigates the use of **software** to translate text or speech from one language to another.
- Automatically convert one natural language into another, preserving the meaning of the input text, and producing fluent text in the output language

# Challenges in Machine Translation

## Lexical Ambiguity

(Example from Dorr et. al, 1999)

book the flight



reservar

(English)

(Spanish)

read the book



libro

# Challenges in Machine Translation Cont.d

## Differing word orders

(Example from Dorr et. al, 1999)

- English word order                           **subject – verb – object**
- Japanese word order                           **subject – object – verb**

IBM bought Lotus (English)

IBM Lotus bought (Japanese)

# Challenges in Machine Translation Cont.d

## Syntactic Structure is not Preserved Across Translations

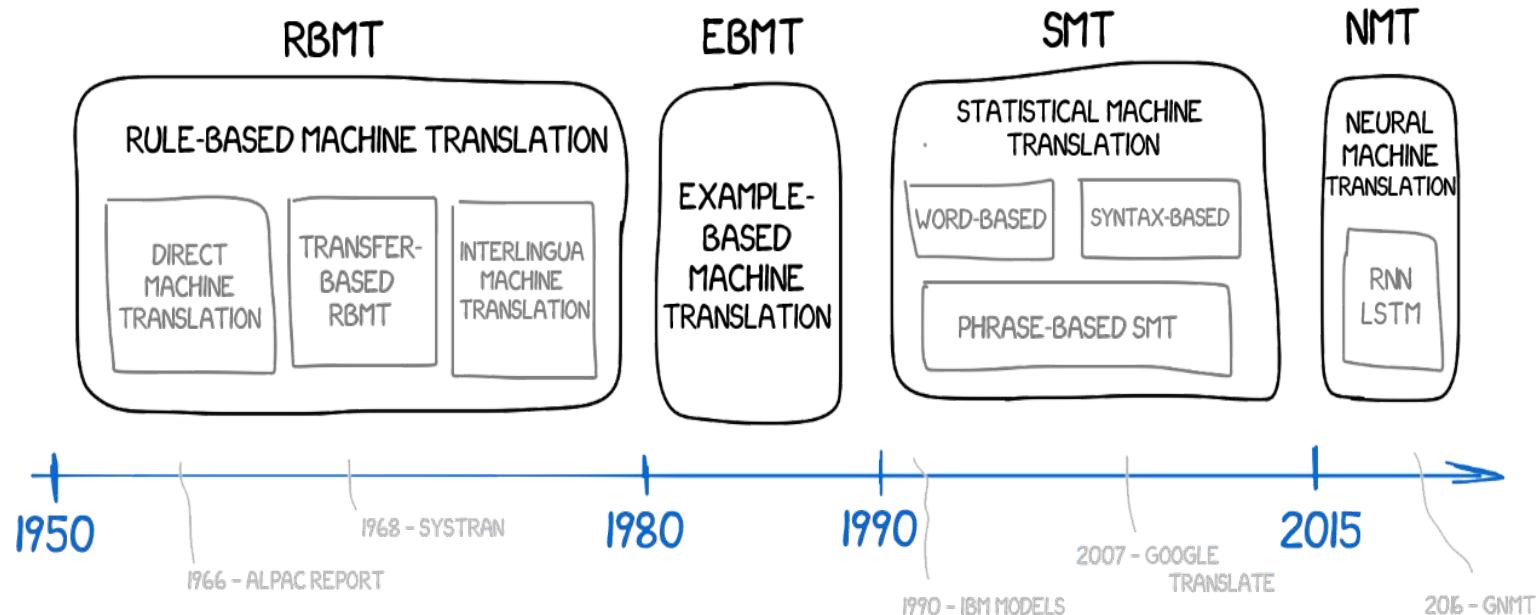
(Example from Dorr et. al, 1999)

The bottle floated into the cave

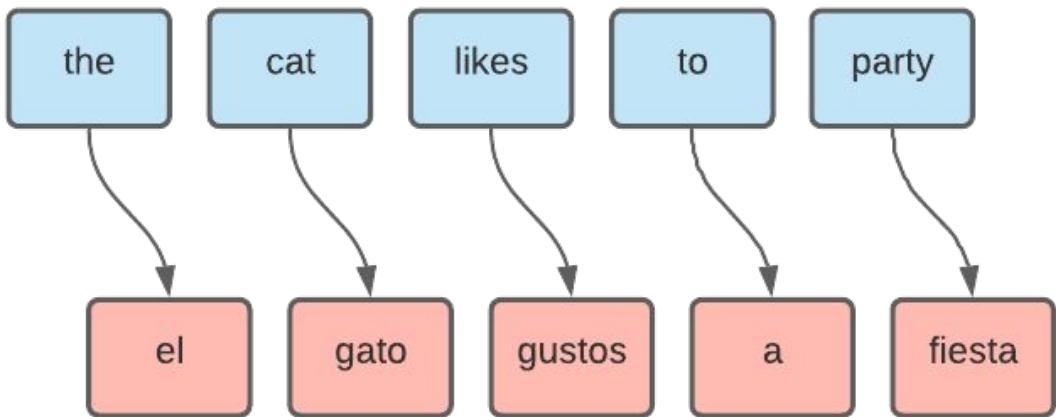


La botella entro a la cuerva flotando  
(the bottle entered the cave floating)

## A BRIEF HISTORY OF MACHINE TRANSLATION



# Rule Based Machine Translation



- Based on **linguistic information** about source and target languages
- Rules are written with linguistic knowledge gathered from linguists

## Challenges

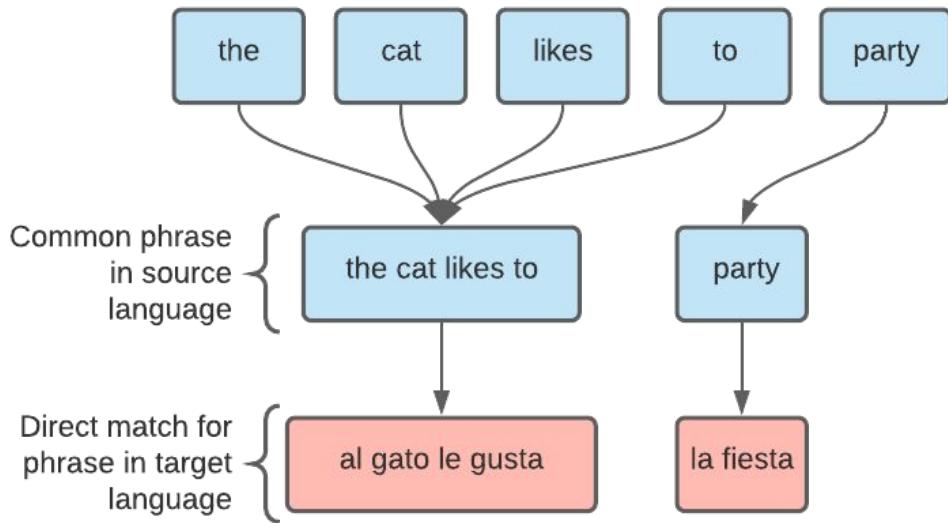
- Requires knowledge and relatively expert human labour
- Many rules can and must be added to improve quality, leading to a very complex system

3.

# Statistical Machine Translation



# Statistical Machine Translation



- Translations are generated on the basis of statistical models whose parameters are derived from the analysis of bilingual text corpora

## Challenges

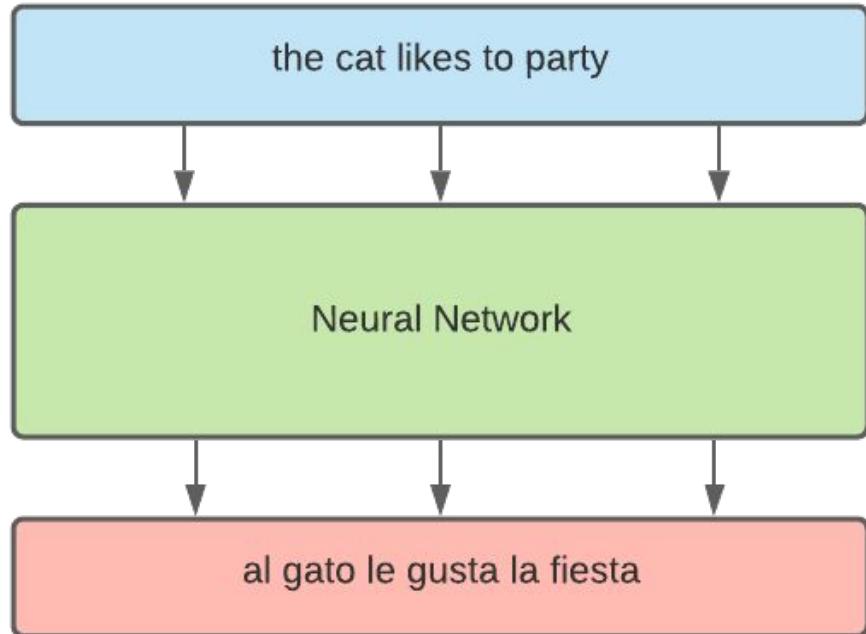
- More memory and time
- Translating material that is not similar to content from the training corpora the accuracy fails

4.



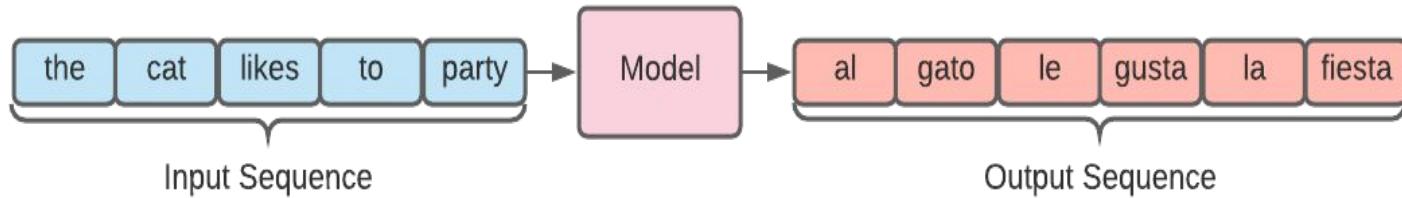
# Neural Machine Translation

# Neural Machine Translation (NMT)



- Task of converting a **sequence of words** from a source language, like English, to a sequence of words to a target language like Spanish or French using **deep neural networks**
- NMT is based on the model of neural networks in the human brain, with information being sent to different “layers” to be processed before output

# Sequence-to-Sequence (seq2seq) models



- Maps a source sequence to target sequence
- Source sequence is the input language
- Target sequence is the output language
- A typical seq2seq model has 2 major components
  - a) an Encoder
  - b) a Decoder

# Transforming Data

the	0
cat	1
likes	2
to	3
party	4
<SOS>	5
<EOS>	6

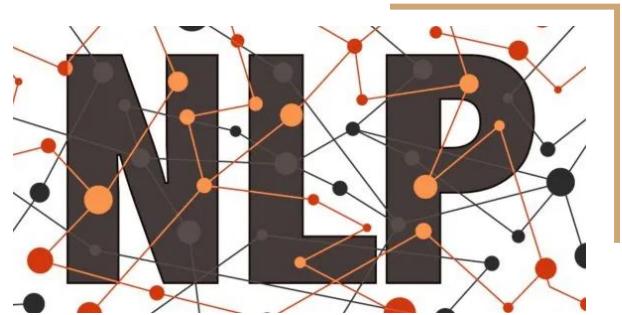
- Textual data is converted into a numeric form
- Each word is transformed into a **One Hot Encoding vector** which can then be inputted into the model
- Assign an index to each unique word in the input language
- In this way, each word has a distinct One Hot Encoding vector and thus we can represent every word in our dataset with a numerical representation

	$\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$
the	cat	likes	to	party	

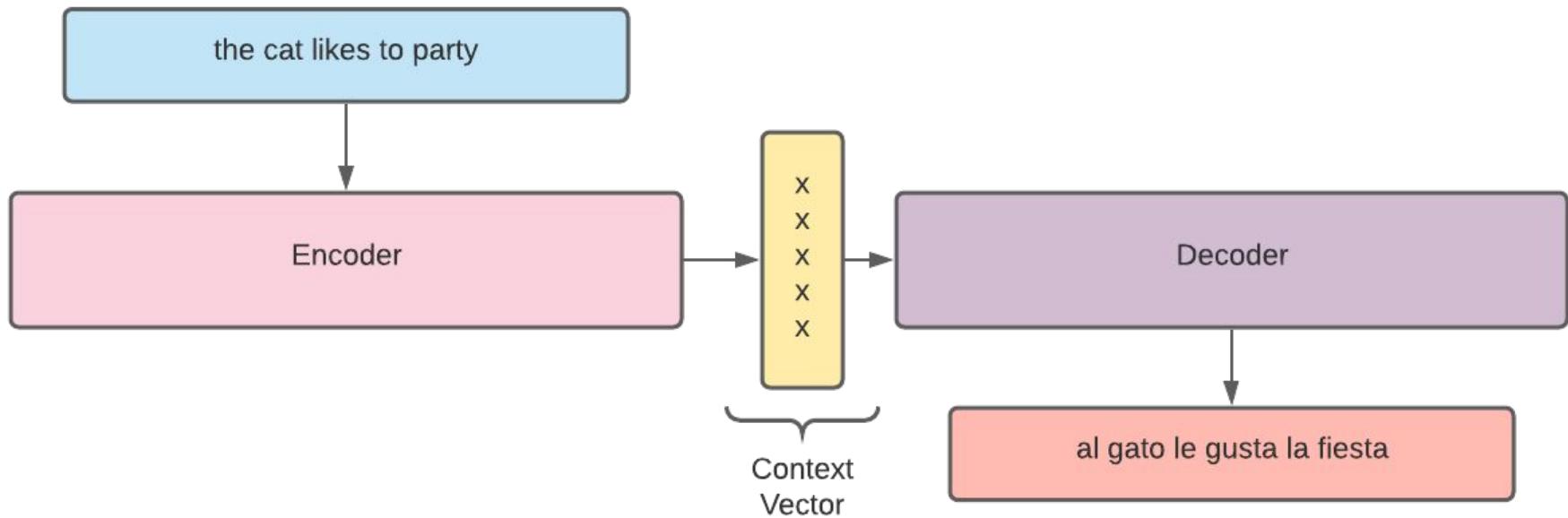
- Each word becomes a vector of length 7 (which is the size of our vocabulary)
- Consists entirely of 0s except for a 1 at the index that was assigned to that word in the before table.

5.

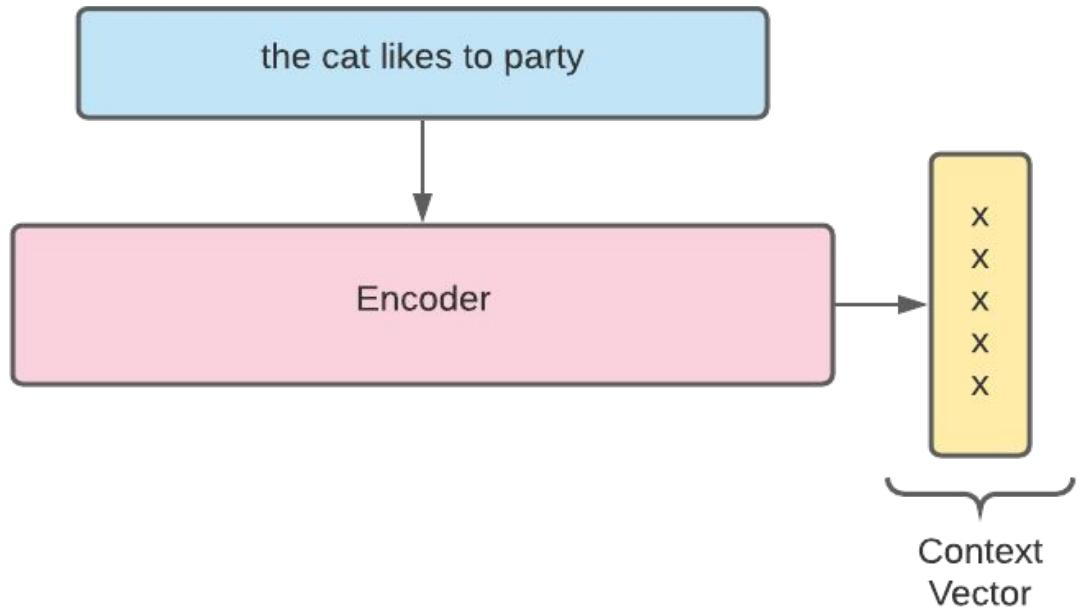
# Encoder Decoder Model



# Encoder-Decoder Models Overview

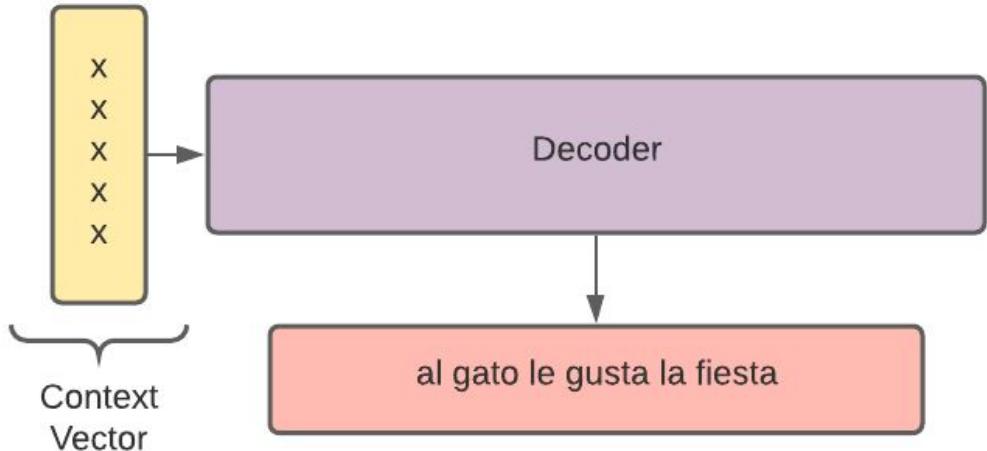


# The Encoder

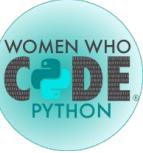


- The encoder creates a single vector which encodes the “meaning” of the input sequence into a single vector.
- This vector is called the ‘context vector’ or ‘thought vector’.

# The Decoder



- The input to the decoder is the context vector created by the encoder.
- The decoder takes that encoded vector and uses it to produce a translation in the target language.



WOMEN WHO

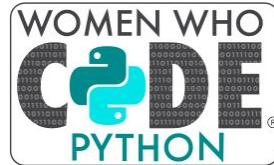
Questions?

R

6

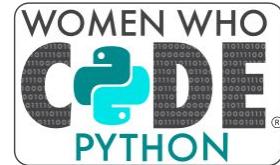
# Google Colab Coding!

<https://colab.research.google.com/drive/1lFIQZvxAttHAnSZQbQY582cnx44WOfKo?usp=sharing>



7

# Recap & Next Steps

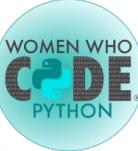


# Recap

In this session, we learnt about:

1. What is Machine Translation
2. Brief history of MT
3. Different types of MT
4. Sequence to Sequence Modelling
5. Encoder Decoder Model
6. For further studies and getting into their details, there are some of the links provided in the Resources slide.

# Upcoming Events!



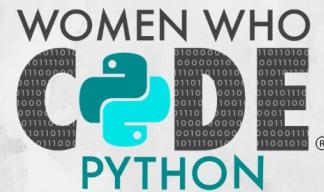
MON  
28  
DEC

Ask Me Anything with Mikiko Bazeley(Machine learning engineer)

9:00 PM – 10:00 PM (EST) | 🔍 Zoom

Register

# Stay Connected

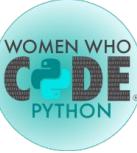


JOIN US ON SOCIAL MEDIA!



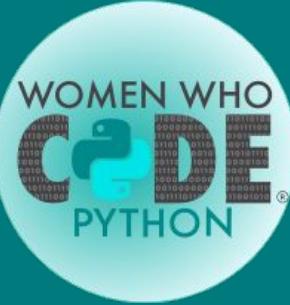
@WWCODEPYTHON

[WOMENWHOCODE.COM/PYTHON](http://WOMENWHOCODE.COM/PYTHON)



# Questions?

Join our Slack channel: #discover-nlp-with-python



Thanks  
everyone!