



# Processamento de Linguagem Natural

João Paulo Aires

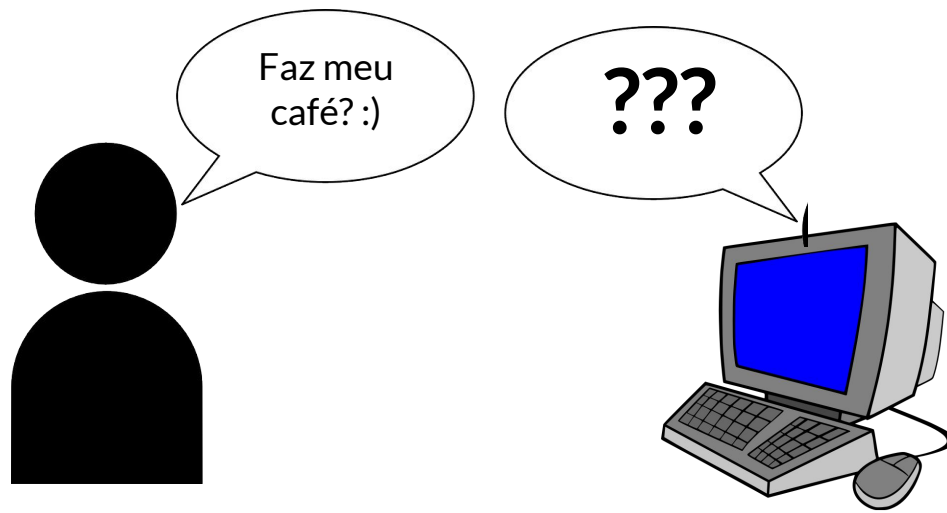
# Linguagem Natural

- **Linguagem Natural** é a língua que nós, seres humanos, utilizamos para comunicação;
- Escrita ou falada, independente do idioma;



# Linguagem Natural

- Máquinas não entendem Linguagem Natural
- Linguagem Natural não é estruturada.





## Processamento de Linguagem Natural (PLN)

- Precisamos processar o texto para extrairmos informações que permitam o entendimento por parte da máquina.





# Processamento de Linguagem Natural (PLN)

João estuda na PUCRS.



# Processamento de Linguagem Natural (PLN)

## Tokenização

['João', 'estuda', 'na', 'PUCRS', '.']



# Processamento de Linguagem Natural (PLN)

## Part-of-Speech

['João/**NNP**', 'estuda/**VB**', 'na/**PREP**', 'PUCRS/**NN**', './**PUNCT**']



# Processamento de Linguagem Natural (PLN)

## Reconhecimento de Entidades Nomeadas

[‘João/**NNP**’, ‘estuda/**VB**’, ‘na/**PREP**’, ‘PUCRS/**NN**’, ‘./**PUNCT**’]

PESSOA

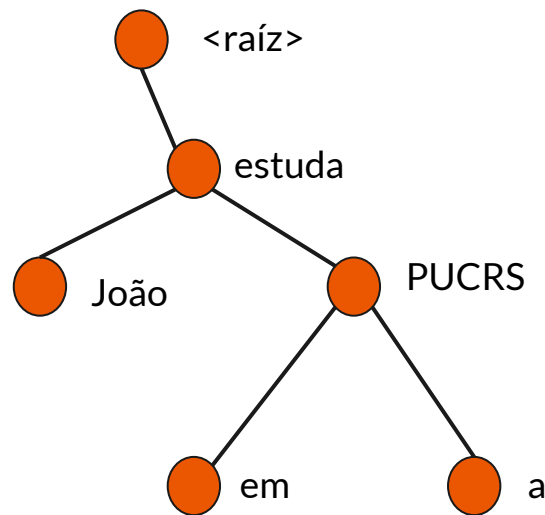
INSTITUIÇÃO  
/EMPRESA



# Processamento de Linguagem Natural (PLN)

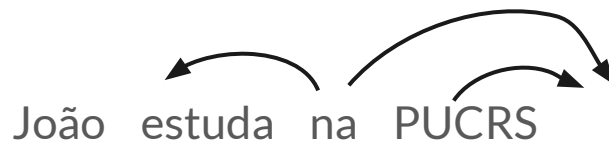
## Árvore Sintática

João estuda na PUCRS.



# Processamento de Linguagem Natural (PLN)

## Árvore de Dependência





# Processamento de Linguagem Natural (PLN)

## Resolução de Co-referência

João estuda na PUCRS, onde ele trabalha.



# Processamento de Linguagem Natural (PLN)

## Resolução de Co-referência

**João** estuda na PUCRS, onde ele trabalha.



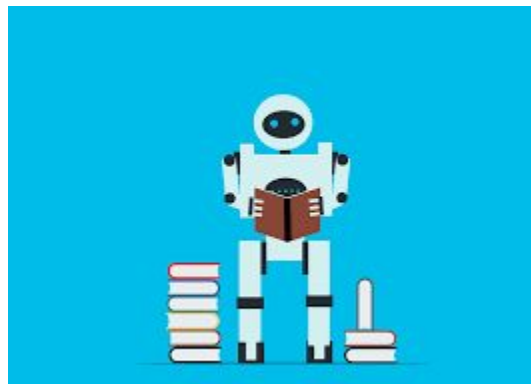
# Processamento de Linguagem Natural (PLN)

## Resolução de Co-referência

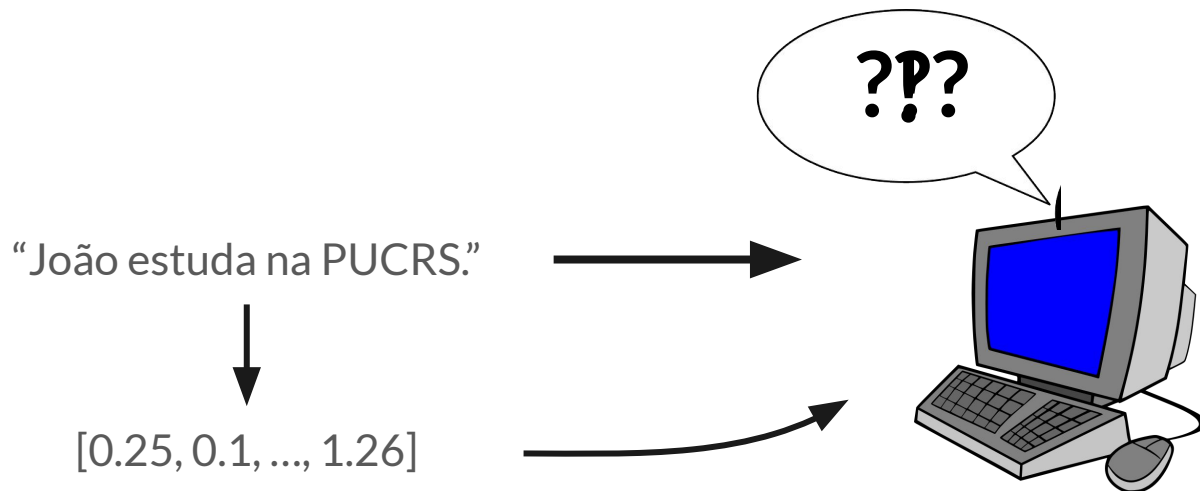


João estuda na PUCRS, onde ele trabalha.

# PLN + Aprendizado de Máquina

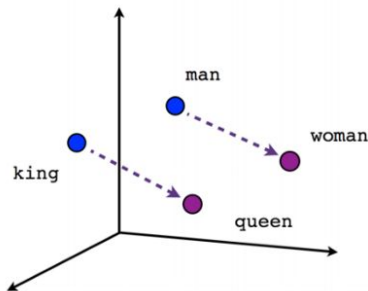


## PLN + Aprendizado de Máquina

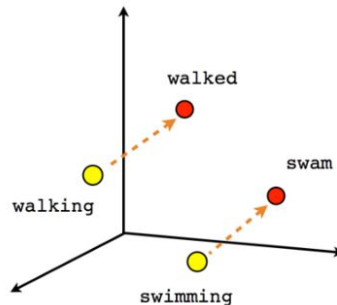


# Representando a Linguagem Natural

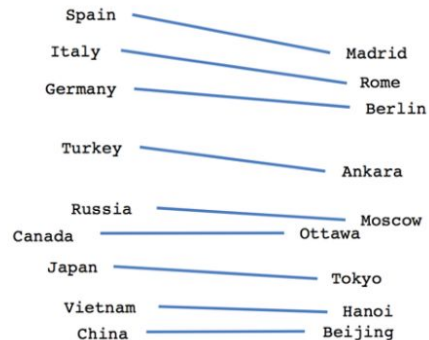
- Os chamados “embeddings” são representações vetoriais com valores densos.



Male-Female



Verb tense



Country-Capital



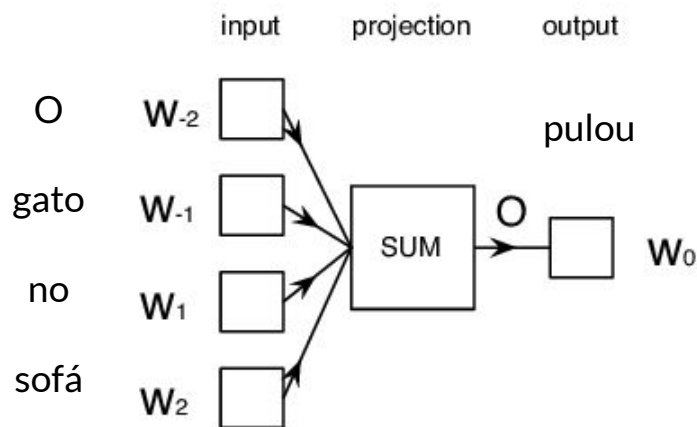


# Representando a Linguagem Natural

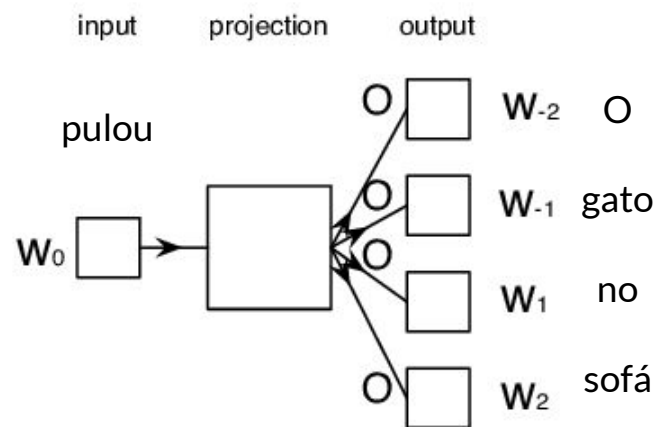
Abram o Jupyter Notebook

# Word2Vec

## CBOW

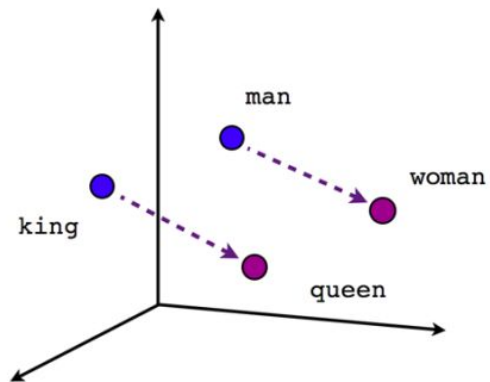


## Skip-Ngram

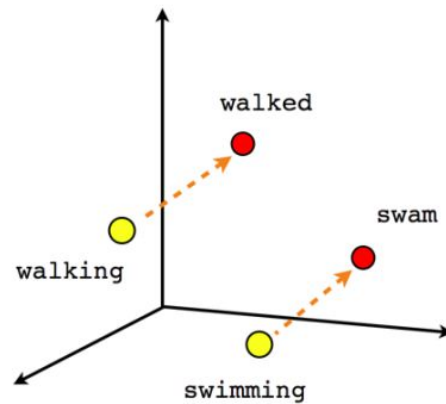




# Word2Vec



Male-Female



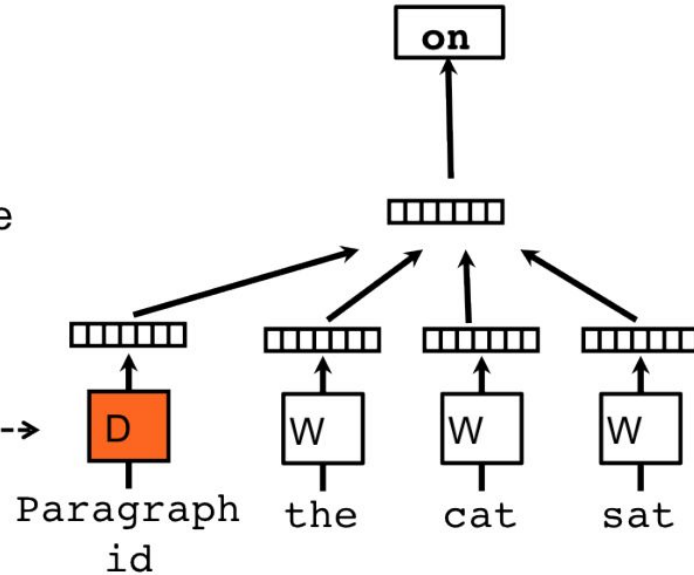
Verb tense

# Sent2Vec

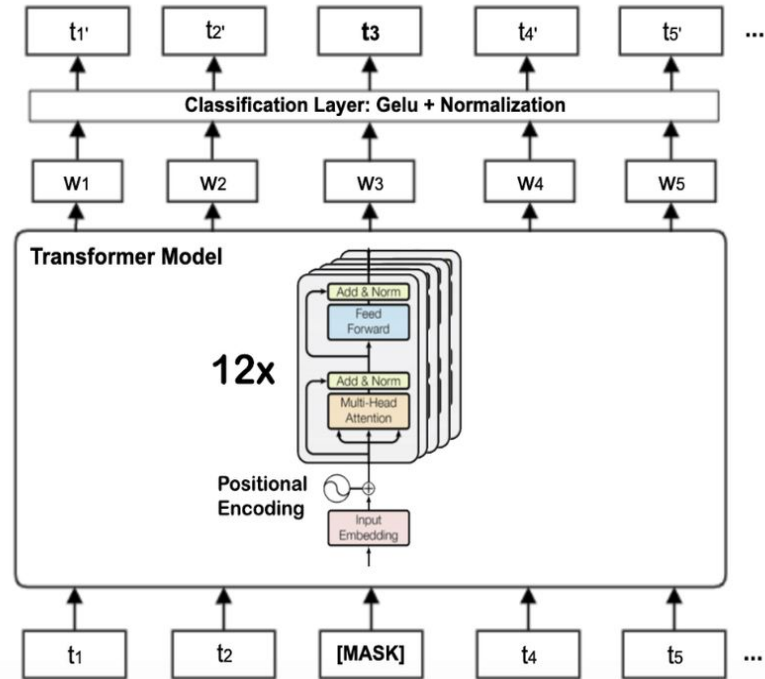
Classifier

Average/Concatenate

Paragraph Matrix----->

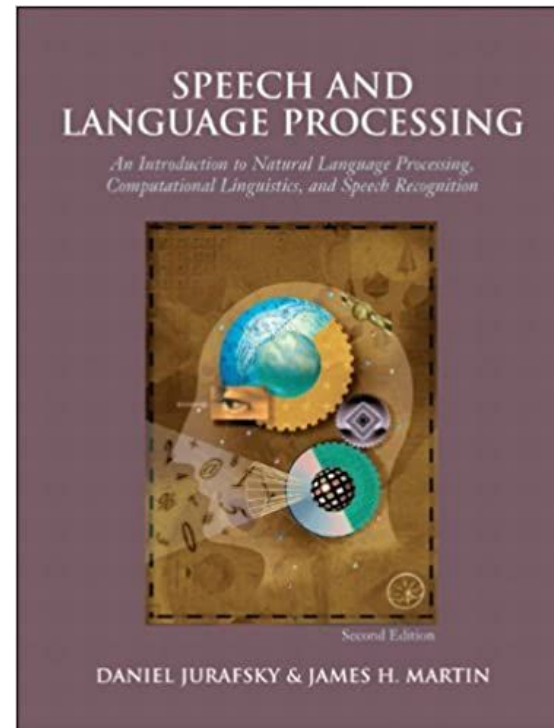


# BERT



# Aprofundamento

- NLTK Book: <https://www.nltk.org/book/>
- Speech and Language Processing: <https://web.stanford.edu/~jurafsky/slp3/>





## Conclusão

- Processamento de Linguagem Natural é uma área cheia de desafios;
- A representação da linguagem natural nos ajuda a extrair informações sintáticas e semânticas;
- O treinamento de modelos para as mais diversas tarefas passa pela definição de uma representação coerente da linguagem natural.



# Obrigado!

João Paulo Aires

[joao.souza91@edu.pucrs.br](mailto:joao.souza91@edu.pucrs.br)