

# TRABALHO

## TRANSFORMER PARA ÁREA DE LINGUAGEM NATURAL

### GERAÇÃO DE CONCLUSÃO DE RELATÓRIO DE AUDITORIA

DISCENTES:

JOSE AUGUSTO AGRIPINO DE OLIVEIRA

JOSÉ EDIVANDRO DE SOUSA JUNIOR

VINICIUS JOSE MIRANDA TOSCANO DE BRITO FILHO

# ARTIGO

## Artigo

Pimentel, C.H.M.; Pires, T.B., Treinamento e análise de um modelo de tradução automática baseado em Transformer

## Objetivo:

- Realizar a tradução de frases do inglês para o francês.

## MATERIAIS E MÉTODOS

Uso do google tradutor para traduzir frases do inglês para o francês

Uso de uma rede transformer para traduzir frases do inglês para o francês

Comparação das traduções usando o score sacreBLEU

# RESULTADO DO ARTIGO

Após a coleta dos textos, foram computados 134.943 *tokens* e 6.453 *types* presentes nos textos em inglês e 143.285 *tokens* e 8.932 *types* em francês. Ao final foram alinhadas 5.494 frases em inglês-francês. A partir disso, o modelo treinado recebeu o *score* sacreBLEU de 7,6467.

Tabela 2. Comparação dos scores sacreBLEU atribuídos às frases geradas pelo modelo treinado e pelo Google Tradutor.

Chave	Score Modelo Treinado	sacreBLEU Google Tradutor
316	32,5	32,6
384	29,5	58,1
635	71,9	50,7
796	22,8	23,2
852	8,3	6,3
950	20,6	12,6
965	21	63,9
1013	82,4	91,2
1166	7,3	7,1
1377	22	26,3
1390	33	7,8
1399	49,2	11,4
1411	49,8	36,1
1418	23,4	10
1437	5,7	14,1
1451	14,4	9,4
1455	29	24,6
1471	37,5	33,5
1486	41,4	34,8
1520	31,9	27,2
Média	31,7	29,0

Fonte: Elaboração própria.

# TRANSFORMER

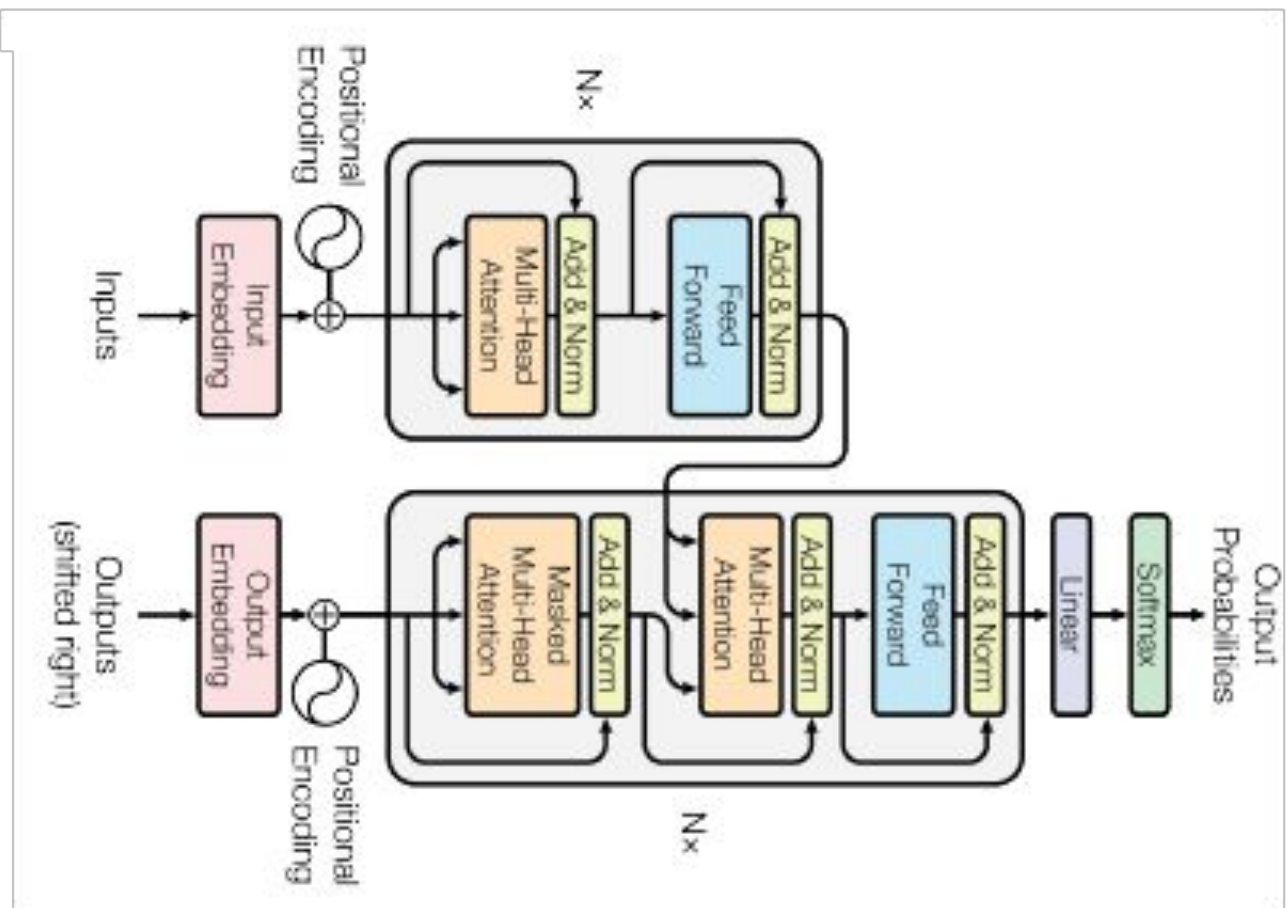


Figure 1: The Transformer - model architecture.

# TRABALHO

## Artigo

Gowtham, R., et al, Samanantar: The Largest Publicly Available Parallel Corpora Collection for 11 Indic Languages

<https://www.youtube.com/watch?v=QCJQG4DuHT0&list=PLTI9hO2Oobd97qfWC40gOSU8C0iu0m2l4>

## Objetivo:

- Realizar tradução do inglês para a língua indiana Kannada.





# MATERIAIS E MÉTODOS

- Dataset com 300000 linhas de textos em inglês e Kannada para ser usado para treinamento e validação
- Uso da arquitetura transformer com dropout
- Inicialização do transformer:
  - `d_model = 512`
  - `batch_size = 30`
  - `ffn_hidden = 2048`
  - `num_heads = 8`
  - `drop_prob = 0.1`
  - `num_layers = 1`
  - `max_sequence_length = 200`
  - `kn_vocab_size = len(kannada_vocabulary)`



# RESULTADO DO TRABALHO

```
[ ] translation = translate("why care about this?")  
    print(translation)
```

⇒ ಏಕೆ ಕಾರಣ ಏನು?<END>

This translated pretty well : "What is the reason. Why" without punctuation.

```
[ ] translation = translate("this is the best thing ever")  
    print(translation)  
    # ಇದು ಎಂದೆಂದಿಗೂ ಉತ್ತಮವಾಗಿದೆ
```

⇒ ಇದು ಅತ್ಯಂತ ಹೊರತಾಗಿದೆ<END>

The translation : "This is very unusual"

# IMPLEMENTAÇÃO

- Códigos:  
[https://drive.google.com/file/d/1oYFzIom72eD3UniGMI5epM\\_d2iNDCZeY/view?usp=sharing](https://drive.google.com/file/d/1oYFzIom72eD3UniGMI5epM_d2iNDCZeY/view?usp=sharing)
- A implementação do transformer encontra-se no arquivo transformer.py
- O arquivo Transformer\_Trainer\_Kannada.ipynb trata do tratamento dos dados, da inicialização do transformer, do treinamento e teste.

# REFERÊNCIAS

## Referências

- Ashish, V., et al, Attention Is All You Need
- Pimentel, C.H.M.; Pires, T.B., Treinamento e análise de um modelo de tradução automática baseado em Transformer
- Gowtham, R., et al, Samanantar: The Largest Publicly Available Parallel Corpora Collection for 11 Indic Languages
- CodeEmporium  
(<https://www.youtube.com/watch?v=QCJQG4DuHT0&list=PLTI9hO2Oobd97qfWC40gOSU8C0iu0m2l4>)