

# Assistente de tradução para gírias brasileiras

1<sup>st</sup> Andre Cruz

1912130026

Centro Universitario IESB

Brasilia, DF

andre.cruz@iesb.edu.br

2<sup>nd</sup> Leonardo Oliveira

1822130012

Centro Universitario IESB

Brasilia, DF

leonardo.oliveira@iesb.edu.br

3<sup>rd</sup> Mariana Cruz

1712082002

Centro Universitario IESB

Brasilia, DF

mariana.cruz@iesb.edu.br

4<sup>th</sup> Pedro Silva

1712082040

Centro Universitario IESB

Brasilia, DF

pedro.sergio@iesb.edu.br

**Abstract—Resumo –** A globalização e o crescimento das redes sociais impactos e evoluiu a linguagem inesperadamente de modo a deixar qualquer um que não vem acompanhando tal processo com afincos totalmente perdido e isolado, sem conseguir compreender palavras que não existiam nos últimos meses. Pensando nisso, surgiu a ideia de desenvolver o Pingu Slangs, com o objetivo de auxiliar o usuário na comunicação do dia a dia que não estão Pingu com o regionalismo do local que estão inseridos e promover o entendimento claro do significado das gírias brasileiras para não somente a aqueles que já conhecem o, porém aqueles que também estão aprendendo a linguagem.

## I. INTRODUÇÃO

O advento da pandemia SARS-Covid levou ao isolamento da população mundial por razões sanitárias, então a introdução do trabalho remoto por diversos setores da indústria no mercado de trabalho foi a melhor solução implementada para impedir que a sociedade estagnasse sem alto risco de contaminação. Isto fez com que a internet torna-se a principal ferramenta para se manter comunicações essenciais e manter a troca de informações diárias, o que consequentemente, gerou grande inserção social no contexto digital fazendo com que houvesse uma maior globalização cultural. Porém o mesmo evento também serviu para expor um problema que, devido ao imenso território brasileiro, o mesmo objeto ou situação é referido de diferentes maneiras dependendo do estado em qual você está inserido, o que torna a comunicação desnecessariamente confusa e complicada quando se há cada vez mais a necessidade da mesma ser limpa, simples e direta.

## II. CONTEXTUALIZAÇÃO

O regionalismo excessivo e a criação de gírias virtuais podem vir a atrapalhar e impedir uma comunicação clara e concisa entre todos os indivíduos inseridos no meio, o que muitas das vezes cria grupos isolados de conversação como aponta o estudo de Preti (1984) sobre gírias que a mesma é caracterizada como um vocabulário especial, sendo considerada um signo de grupo, a princípio secreto, de domínio exclusivo de uma comunidade social restrita. Portanto como forma de diminuir o problema distanciamento social virtual, optamos pela criação de um aplicativo que irá primeiramente coletar dados de uma rede social popular, como exemplo "Twitter" e realizar uma tradução do contexto para um português mais acessível, com o intuito de ensinar e inserir o usuário em um ciberespaço que evolui constantemente.

Acoplado com uma interface de fácil entendimento e uma funcionalidade de tradução de gírias virtuais e regionalismos que facilita a inserção do usuário no contexto moderno, o Pingu Slangs busca trazer uma maior unificação do linguajar brasileiro de maneira intuitiva e simples, fazendo com que o novo aprendizado flua de maneira leve e divertida.

Com a funcionalidade de tradução de gírias implementada de maneira efetiva, é possível olhar as demais possíveis aplicações deste modelo de projeto, em especial a tradução de jargões e termos científicos nos campos da medicina e da legislação, facilitando o aprendizado, e possivelmente até mesmo tendo a funcionalidade de traduzir tais termos para outros idiomas, impactando positivamente esses dois campos de estudos e promovendo parcerias de estudos internacionais entre pesquisadores e escolares.

## III. PROBLEMA

Esse Trabalho tem como justificativa apresentar um aplicativo do qual ajudará o máximo de brasileiros possíveis, ou usuários buscando aperfeiçoar o seu português, a entender o contexto de uma mensagem que possua gíria de modo a não serem apartados de um grupo devido a não possuírem a mesma identidade cultural.

Dessa forma, o presente estudo tem a proposta de responder a seguinte pergunta: **Como facilitar o entendimento de frases que possuem gírias afim de compreender todo o contexto da mensagem?**

## IV. OBJETIVO GERAL

Foi idealizada para o desenvolvimento deste projeto uma aplicação gratuita e de simples utilização que proporcione a tradução de gírias de uma maneira intuitiva que possa ser utilizada e compreendida por indivíduos de todas as faixas etárias e de todas as regiões do país.

## V. OBJETIVO ESPECIFICO

- Estudar e analisar, em cima do tema proposto, como programar um aplicativo que facilite e contribua para o processo de entendimento de gírias brasileiras;
- Criar o design do aplicativo, de forma simples e intuitiva na ferramenta FIGMA;
- Criar um sistema de autenticação do aplicativo.
- Treinar e inserir um chatbot na aplicação para auxiliar o usuário

- Preparar datasets com as principais gírias utilizadas online para pesquisa e identificação facilitada e rápida
- Implementação da API do Google Maps no aplicativo

## VI. REFERENCIAL TEORICO

Este estudo teve como principal contribuição a reflexão sobre o uso de dispositivos móveis como ferramenta para tradução de gírias brasileiras, pretendendo auxiliar no processo de entendimento da frase ou texto, principalmente nas redes sociais.

Percebendo que, atualmente, a população está conectada por meio, principalmente, de celulares, é necessário analisar e traduzir as principais gírias que circulam nas redes sociais, para que o máximo de pessoas consiga entender o que está sendo transmitido na mensagem.

## VII. TRABALHOS CORRELATOS

- Dialog Flow - BERT (Bidirectional Encoder Representations from Transformers), um artigo publicado por Google AI Language - Twitter Slang, dataset

## VIII. METODOLOGIA

Para o processamento de linguagem natural (NLP) foi optado por BERT (Bidirectional Encoder Representations from Transformers), um artigo publicado recentemente por pesquisadores pertencentes a Google AI Language que demonstrou resultados estupendos em resolver problemas NLP, dentre eles respostas de perguntas (SQuAD v1.1), Interferência de Linguagem Natural, entre outros.

Devido ao fato de BERT usar um Transformador, um mecanismo que aprende relações contextuais entre palavras e subpalavras em um texto, com apenas um codificador se fazendo necessário para a leitura e geração da linguagem, é possível então ler a sequência inteira de palavras de uma vez independente da direção, o que inclui o contexto das palavras, diretamente o oposto dos modelos direcionais tradicionais. Com isso BERT consegue superar as limitações de contexto que modelos direcionais apresentam, sendo assim o modelo mais indicado para a tarefa designada neste projeto.

Utilizando a localização do usuário fornecida pela API Maps da Google, é possível então traçar uma relação entre sua posição e as gírias mais predominantes na área. Essa informação é crucial para poder limitar e filtrar um conjunto de opções a serem permeadas por BERT de modo a otimizar processamento a ser feito e possivelmente exibir um grupo de escolhas inicial para o mesmo no menor tempo de execução possível.

O Twitter Slang tornou-se o dataset padrão para a busca de parâmetros devido a atual falta de um dataset com o foco na língua brasileira. Para suplementar os dados será feito o uso da API fornecida pela twitter em que é possível buscar e classificar tweets baseados em localização e metadados que serão usados para alimentar a base de dados do BERT.

Por fim, para auxílio a usuários novos a tecnologia e aqueles que necessita de suporte técnico, iremos utilizar o chat bot fornecido pela Google a fim de complementar e facilitar a experiência do mesmo.

## IX. RESULTADOS

- Com o objetivo de desenvolver um aplicativo para facilitar o entendimento das gírias utilizadas no Brasil, utilizamos um web crawler, para realizar a busca das gírias nas redes sociais e realizar a tradução;

- Desenvolver inicialmente o desing das telas do aplicativo para facilitar e identificar o que deve ser desenvolvido em cada etapa;

- Realizar a autenticação dos usuários, para que estes tenham acesso a página de tradução das gírias.

## X. TRABALHOS FUTUROS

O estudo futuro e continuidade de pesquisa nesta área é altamente recomendado devido as possíveis aplicações de um identificador e decodificador de linguagem natural voltado a um linguajar específico com aplicações sociais amplas, como por exemplo, no campo da medicina aonde ossos do corpo humano possuem nomes diferentes variantes com a nacionalidade do usuário. Este estudo então, poderá ser uma ferramenta poderosa para a integração e evolução da medicina moderna com parcerias não mais limitadas por barreiras linguísticas. Outra possível exemplo de campo que se beneficia por tal estudo é o judiciário que por possuir linguajar coloquial e jargões técnicos de extrema complexidade, muita das vezes derivados do latim e adaptados conforme regionalidade, acaba com grandes intervalos impostos por limitações de comunicação.

## XI. CONCLUSÃO

Este trabalho possibilitou entender como podemos diminuir o distanciamento social virtual, causado, principalmente, pelo uso excessivo de gírias e a necessidade de ferramentas que façam a integração de gírias com o português coloquial quando possível. Para se atingir uma compreensão dessa realidade, foi idealizada uma aplicação simples e gratuita que proporciona a tradução de gírias do Twitter de forma intuitiva, podendo ser utilizada por indivíduos de todas as regiões e faixa etárias. Este trabalho tem o intuito social de ajudar o máximo possível de habitantes no Brasil e melhorar o entendimento de uma mensagem que, devido a alguma interferência, acaba não sendo compreendida de forma clara e coesa.

## XII. BIBLIOGRAFIA

- Silva, Carlos Patriota, Luciene. A comunicação na era digital: as gírias das redes sociais. Revista de Letras JUÇARA, Caxias – Maranhão. 2020.
- Arraz, Fenando. Gíria socioeducativa: recurso linguístico utilizado pelos adolescentes e/ou jovens que cumprem medida de internação socioeducativa. PUC/MG. 2019.
- Galindo, Marilene. A gíria como processo comunicativo das reeducandas do presídio Santa Luzia: construção de um glossário. UFAL. 2021.
- Jacob, Devlin Chang, Ming-Wei Lee, Kenton Toutanova, Kristina. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv, 2019.
- VASWANI, Ashish et al. Attention is all you need. Advances in neural information processing systems, v. 30, 2017.
- Araújo, Gisele Santos, Marcus Roberto de

Araújo, and Maria do Carmo Acácio de Sousa. "Gírias, códigos linguísticos como afirmação e identidade de um grupo: uma análise e reflexão da possibilidade de uso no ensino e aprendizagem da língua portuguesa." *EntreLetras* 7.1 (2016)