

# GíriasIO: Aproximando Gerações, Grupos e Dialeto

Matheus Pedra Puime Feijó<sup>1</sup>, Rodrigo Oliveira<sup>1</sup>

<sup>1</sup>Programa de Pós-Graduação em Informática – Universidade Federal do Rio de Janeiro (UFRJ)

Caixa Postal 68.530 – Cidade Universitária – RJ – Brazil

{matheuspuime, rodrigo.santosprogressao}@gmail.com

**Abstract.** *The use of slangs can be considered as a facilitator between the communication of individuals in the same group, however, it has the potential to make it difficult to establish relationships between outsiders. This article presents a slang translation tool in order to improve the dialogue between two individuals from different linguistic niches. The proposed approach identifies slangs in messages of an instant messaging service, which can be altered by their synonyms, and may facilitate the understanding of excluded groups.*

**Resumo.** *A utilização de gírias pode ser considerada como um facilitador entre a comunicação de indivíduos de um mesmo grupo, porém tem o potencial de dificultar o estabelecimento de relacionamentos entre indivíduos externos a esse grupo. Este artigo apresenta uma ferramenta de tradução de gírias para possibilitar a melhor interlocução entre dois indivíduos de nichos linguísticos diferentes. A abordagem proposta identifica gírias presentes em mensagens de um serviço de mensagens instantâneas, no qual podem ser alteradas pelos seus sinônimos, podendo facilitar o entendimento de grupos excluídos.*

## 1. Introdução

O surgimento de novas tecnologias e o aprimoramento dos dispositivos móveis impulsiona um amplo crescimento e estreitamento das relações interpessoais, principalmente a partir da utilização de aplicativos de mensagens de textos, assim criando lugares de comunicação como é citado no trabalho de Nouwens, Griggio e Mackay (2017). Esses lugares oferecem ao usuário a criação de um ecossistema em que podem interagir com a sua rede de contatos, utilizando suas características pessoais.

Nesses espaços, a utilização de gírias é frequente, sendo ela uma manifestação da língua que ocorre, principalmente, por meio do modo oral em contextos informais e em situações nas quais os indivíduos interagem em grupos com seus pares, reforçando a ideia de pertencimento, inclusive, identidade dos participantes com o grupo e vice-versa [Eble 1996]. Contudo, indivíduos não pertencentes ao grupo possuem dificuldade de entendimento e, conseqüentemente, a comunicação pode ser prejudicada, dessa forma tendo o potencial de impossibilitar relações entre o grupo e o indivíduo não pertencente.

Para mitigar a barreira de comunicação entre as duas partes, é necessário que ambas estabeleçam um meio de colaboração com o intuito de que as gírias utilizadas tenham significados entre os grupos. Apesar do amplo desenvolvimento de sistemas de tradução por máquina de diferentes idiomas, ainda há pouco foco na variabilidade linguística de uma própria língua. Com isso, este artigo propõe uma formulação inicial de uma aplicação para tradução de gírias buscando auxiliar a interação de grupos

linguísticos diferenciados juntamente com uma análise de sentimento dessas gírias, para auxiliar na compreensão contextual onde ela se insere. O intuito desta aplicação é ser utilizada para várias finalidades, como transcrever um diálogo com gírias para o padrão mais formal, auxiliar na compreensão de textos com ampla variedade linguística e, ainda, diminuir a distância de comunicação entre indivíduos de classes linguísticas distintas.

Além desta introdução este trabalho está organizado em mais oito seções. As seções 2 e 3 apresentam noções introdutórias a respeito dos temas de gírias e tradução. Na seção 4 são expostos os trabalhos correlatos. A seção 5 é apresentada a metodologia definida para solução da problemática encontrada. Na seção 6 é apresentado a proposta deste referido trabalho. A seção 7 apresenta a proposta de uma futura avaliação e validação da ferramenta. Por fim, a seção 8 apresenta as considerações finais.

## **2. Gírias**

O campo de estudo que se concentra na pesquisa da variação de uma língua por vários fatores externos é a sociolinguística, no qual a língua e sua variedade são consideradas como entidades inseparáveis, tão essenciais à existência uma da outra [Preti 1984]. Diversas teorias dentro da própria sociolinguística buscam explicar essa variabilidade e seus fatores e impactos, incluindo Labov (2008) da Teoria da Variação, que compreende essas mudanças como manifestações no momento de sua expressão, com influência de inúmeras variáveis linguísticas ou sociais, como as condições do falante, sua participação em um grupo, a situação onde a língua é utilizada, ou todos ao mesmo tempo.

Para Preti (1994), esses fatores de variação extralinguísticas são agrupados em três grandes classificações: (i) Geográficas, sobre variações regionais e de localidade (ii) Sociológicas, nos casos de variações provenientes da idade, sexo, profissão entre outros fatores e (iii) Contextuais, todas as diferenças na linguagem que um locutor pode utilizar por influências externas, como, por exemplo, o assunto, o tipo de ouvinte, ou lugar em que o diálogo ocorreu. Para Eriksen (2010) a noção de gíria transmite a informalidade em um diálogo se opondo a norma culta da língua utilizada. Logo, o contexto e uso social de uma palavra que irá determinar o seu reconhecimento como uma gíria em si. Essa pesquisa busca desenvolver um ambiente no qual grupos linguísticos diferentes de um mesmo idioma possam de forma colaborativa fornecer e identificar gírias, auxiliando na tradução para termos comuns a ambos os grupos compreenderem os termos nativos do grupo oposto.

## **3. Tradução**

Segundo o dicionário Michaelis da língua portuguesa, tradução é a técnica que consiste em traduzir uma palavra, enunciado, texto ou obra de uma língua para outra, possibilitando sua compreensão por alguém que não domina a língua em que o enunciado foi emitido. Entretanto, encontramos definições que abordam essa questão de forma mais ampla, apresentando a tradução além do usual, de ressignificar palavras em línguas distintas, mas também reformular os símbolos verbais dentro de uma própria língua, chamada de tradução intralingual que utiliza outra palavra, mais ou menos sinônima, ou a uma sequência de palavras ou locuções para transcrever em lugar do termo original [Jakobson 1969]. Em palavras de um idioma que não se conhece o significado, uma simples consulta ao dicionário é suficiente, porém com as gírias,

encontrar um conjunto de significados atualizado que atendam a todos os contextos e variações é difícil e custoso, manter um sistema apoiado em colaboração permite que os próprios participantes das conversas possam manter os significados reais ao longo do tempo e novos termos em uso atual.

#### **4. Trabalhos Correlatos**

O trabalho de Galván (2015) propõe a criação de uma ferramenta de análise com o objetivo de tradução de gírias entre os idiomas inglês e espanhol, fornecendo um termo similar no outro idioma, para isso foi utilizado o Fillmore's Frame Semantics Theory, que divide os conceitos analisados em diferentes termos e engloba em classificações diferentes a partir da percepção humana com isso auxiliando na tradução. Outro estudo relevante é o de Zhou e Fan (2013), no qual é analisado a questão social de surgimentos de gírias nos Estados Unidos, com isso é possível entender melhor sobre o público alvo da aplicação desenvolvida utilizando este trabalho como referência.

Adicionalmente, foram encontradas aplicações voltadas ao entendimento e tradução de gírias, como a aplicação de Flames (2019) no qual fornece diversas gírias em inglês e seus significados. Ainda existem outros exemplos como Urban Dictionary (2019), All Slang (2019), Qual É a Gíria (2019), contudo, esses só fornecem ao usuário buscas de gírias com retorno de seu significado. Logo, não foram encontrados trabalhos voltados a realização de tradução de mensagens contendo gírias no idioma português, assim, a aplicação a ser desenvolvida nesta pesquisa poderá ser considerada como uma inovação na área de análise e tradução de textos.

#### **5. Metodologia**

Com o presente trabalho, pretende-se ampliar o auxílio no entendimento de gírias, através de uma solução para dispositivos móveis onde os próprios usuários possam inserir novas gírias ao sistema. Foi realizado um levantamento da literatura sobre análise e tradução de gírias para compreender as características e público alvo da pesquisa. Identificando artigos relevantes como o de Rocha, Pimentel e Diniz (2014) com o intuito de reconhecer as atitudes de um indivíduo durante sua atuação em um grupo de troca de mensagens, identificando questões como o tempo de reflexão de um participante do grupo da leitura até o entendimento da mensagem, além de outros já apresentados na seção de trabalhos relacionados, contribuindo para o melhor entendimento do contexto e desenvolvimento da solução.

O passo seguinte foi um levantamento dos requisitos e funcionalidades envolvendo a solução a ser desenvolvida, como a plataforma a ser utilizada pelos usuários, sendo escolhido a rede de troca de mensagens Telegram que possui uma alta quantidade de usuários [Smith 2019] e além de aceitar a inclusão de códigos ao seu sistema a partir de Web Robôs. Para a definição de coleta de dados das gírias e seus respectivos significados, foi determinado a criação de um formulário para captação das gírias mais relevantes e que não possuem um reconhecimento expressivo da população. É importante destacar que essa demonstração inicial restringe-se, exclusivamente, a gírias de residentes do estado do Rio de Janeiro com o intuito de realizar um controle inicial dos dados de entrada.

A partir disso, foram definidos os módulos de desenvolvimento para um esboço da aplicação e seus cenários e casos de uso para modelar o comportamento de

identificação de gírias em um diálogo entre um grupo de indivíduos, como suportar um critério colaborativo para a inserção de novas gírias por meio dos usuários.

## 6. Mecanismo de Tradução de Gírias: GíriasIO

O GíriasIO apresenta uma arquitetura simples (Figura 1), onde utiliza como base a interface do aplicativo de mensagens Telegram, o papel da ferramenta é captar as mensagens e traduzi-las para um padrão mais próximo da linguagem formal do idioma, que a princípio será de entendimento mais fácil para os usuários. O sistema foi adicionado a um Web Robô (@gíriasIO\_bot), que realiza tarefas automáticas, atuando com usuários e obedecendo a comandos específicos, sua implementação pode ser vista na Figura 2(a).

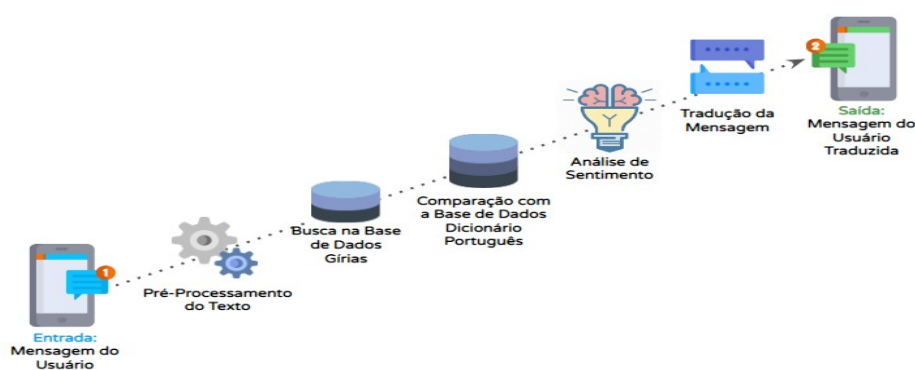


Figura 1 - Arquitetura do Sistema GíriasIO

O sistema foi desenvolvido utilizando a linguagem Python e os recursos da biblioteca NLTK, no qual o seu código pode ser acessado na plataforma GitHub em <http://bit.ly/gíriasIO>. Para processamento das mensagens é realizada uma etapa de pré-processamento, que divide a mensagem em unidades mínimas de texto. É realizada a eliminação das pontuações e exclusão de palavras sem sentido necessário para a tradução, como artigos, preposições e entre outros. Para a detecção de gírias, é realizado a comparação com a base criada a partir do formulário disponibilizado. O catálogo contém 50 gírias iniciais (disponível em <http://bit.ly/Bgírias>) para teste do modelo, de uso comum a falantes do grupo geográfico do Rio de Janeiro, principalmente da Baixada Fluminense, com faixa etária de 18 a 28 anos.

A partir de uma gíria identificada dentro do catálogo, esta é um candidato a sofrer a tradução para um termo semelhante, entretanto, algumas palavras podem ser usadas, ou não, como gíria, principalmente se a mesma também possui um uso formal, que é alterado para uso como gíria, em certo contexto, como por exemplo a palavra droga, que pode ser usada no sentido literal, significando substância química ou pode ser usada com o sentido de algo ruim. Para essa análise, uma base da língua portuguesa retirada do editor de texto LibreOffice foi adicionada ao projeto para comparar os termos potenciais de tradução e detectar se a palavra se encontra tanto na base de dados de gírias como no dicionário formal. Em caso positivo, uma análise de contexto seria necessária, para determinar qual o real sentido de uso na mensagem em questão, porém a ferramentas não o realiza, apresentando ao usuário que a palavra pode ser uma gíria, conforme é mostrado na Figura 2(b). Ao mesmo tempo que a tradução da gíria pelo sistema é realizada, também ocorre uma análise de sentimento de toda a mensagem

coletada, com o intuito de trazer mais informações a respeito do contexto da gíria na mensagem enviada, trazendo mais informações ao usuário do grupo linguístico diferente. Utilizando uma base de treinamento previamente estabelecida é identificado se a mensagem enviada possui um teor negativo ou positivo, contribuindo para um melhor entendimento do usuário.

Outra funcionalidade implementada ao robô é a possibilidade de o usuário colaborar com a inserção de novas gírias como é demonstrado na Figura 2(c). Assim, novos termos com seus significados e sinônimos podem ser adicionados ao sistema depois de serem analisados por um administrador. Além disso, quando uma frase contendo uma gíria é identificada pelo sistema o mesmo insere no banco de dados para aprimorar a análise a partir de seu sentimento. Estas questões acabam por se assemelhar ao Modelo 3C de Colaboração, visto que alcança a comunicação a partir do sistema de troca de mensagens, a cooperação a partir do auxílio do usuário a aprimorar o sistema, e a coordenação com o sistema analisando e identificando as gírias nas mensagens, como é mostrado na separação do Modelo 3C exposto por Pimentel et al. (2006).

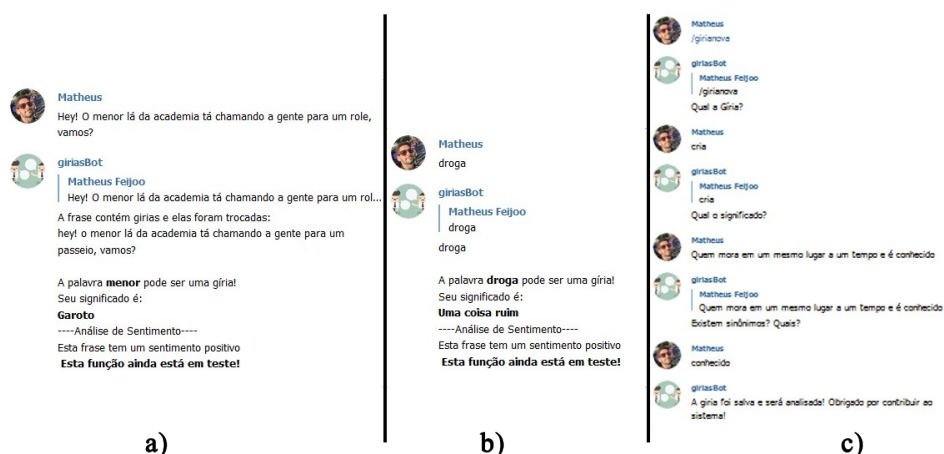


Figura 2 - Exemplo de Utilização do Sistema GíriasIO

## 1. Aspectos de Avaliação

O GíriasIO será avaliado e validado por meio de um cenário de teste a partir de um ambiente controlado e da realização de questionários e entrevistas com indivíduos de diferentes grupos linguísticos, a fim de averiguar a aceitação do sistema e medir o impacto no entendimento com as mensagens traduzidas. Para tal, será utilizado o modelo de Teoria Unificada de Aceitação e Uso da Tecnologia (UTAUT2). O UTAUT2 aplica as variáveis: Intenção de Comportamento e Intenção de Uso [Venkatesh; Morris; Davis, 2003]. Com isso, os dados coletados por meio de questionários e entrevistas serão avaliados utilizando esses dois critérios e, com as informações geradas, será possível identificar os pontos positivos e negativos assim como o nível de aceitação do aplicativo.

## 7. Conclusão

Com o andamento do projeto de pesquisa foi possível identificar a real necessidade do desenvolvimento de aplicações voltadas para a inserção daqueles que se sentem excluídos por conta da variedade linguística de um idioma. Com a utilização do GíriasIO, há a possibilidade de que as barreiras da linguagem sejam quebradas, assim

facilitando a comunicação entre indivíduos de diferentes grupos. A utilização de um sistema já consolidado como o Telegram para ser utilizado como base para o desenvolvimento do tradutor de gírias pode agregar ainda mais adeptos a sua utilização, tendo em vista o alto de usuários ativos na plataforma.

A fim de dar continuidade à aplicação, pretende-se avaliar os resultados do sistema com usuários para medir a eficiência e entendimento da tradução, de maneira que recebe um feedback para otimização da ferramenta, como descrito na seção 7, além de expandir as bases de dados de gírias para abranger mais contextos linguísticos, como de localidades variadas, de contextos diferentes, entre outros, incluindo até mesmo, inferências das características do usuário para identificar quais grupos de gírias são mais prováveis de utilização. Por fim, fornecer à comunidade acadêmica recursos para realização de análise a partir de gírias, visto que será criada uma base com as gírias coletadas.

### Referências Bibliográficas

- AllSlang. (2019). <https://www.allslang.com/>. Acesso: 12 Jun. 2019.
- Eble, C. (1996). *Slang and Sociability*. The North Carolina University Press.
- Eriksen, M. H. (2010). *Translating the use of slang*. Aarhus University.
- Flames, Dev Studios. (2019). *Slang Dictionary*. [bit.ly/arADS](http://bit.ly/arADS). Acesso: 06 jun. 2019.
- Galván, B. (2015). “*Trapicheando*” on Baltimore’s Corners A Semantic Analysis of The Wire’s Slang Terminology and its Translation. *Procedia - Social And Behavioral Sciences*. Elsevier BV.
- Jakobson, R. (1969). *Aspectos linguísticos da tradução*. Linguística e comunicação.
- Labov, William. (2008) *Padrões sociolinguísticos*. Parábola Editorial.
- Pimentel, Mariano et al. (2006). *Modelo 3C de Colaboração para o desenvolvimento de Sistemas Colaborativos*. In Anais do III SBSC.
- Nouwens, M.; Griggio, C.; Mackay, W. (2017) “*WhatsApp is for family; Messenger is for friends*”: *Communication Places in App Ecosystems*. ACM Press.
- Preti, D. (1984). *A gíria e outros temas*. EDUSP.
- Preti, D. (1994). *Um estudo sociolinguístico do diálogo na literatura brasileira*. EDUSP.
- Qual é a gíria?. (2019). <https://www.qualeagiria.com.br/>. Acesso: 12 Jun. 2019.
- Rocha, E.; Pimentel, M.; Diniz, M.. *A relação entre o tamanho do grupo e a produção de mensagens de bate-papo*. In Anais do XI SBSC.
- Smith, Craig. (2019). *10 Telegram Stats and Facts*. [bit.ly/2IP3aLC](http://bit.ly/2IP3aLC). Acesso: 01 ago. 2019.
- Urban Dictionary. (2019). <https://www.urbandictionary.com>. Acesso: 12 Jun. 2019.
- Venkatesh; Morris; Davis. (2003). *User Acceptance of Information Technology: Toward a Unified View*. *Mis Quarterly*, JSTOR.
- Zhou, Yanchun; Fan, Yanhong. (2013). *A Sociolinguistic Study of American Slang. Theory And Practice In Language Studies*. Academy Publication.