

# **LibrasTube:**

## **Um Aplicativo para a Criação de Janelas de Libras Sincronizadas com Vídeos do YouTube**

Leandro Henrique de Souza Santos



CENTRO DE INFORMÁTICA  
UNIVERSIDADE FEDERAL DA PARAÍBA

João Pessoa, PB  
Junho - 2018

Leandro Henrique de Souza Santos

# LibrasTube

Monografia apresentada ao curso Engenharia de Computação do Centro de Informática, da Universidade Federal da Paraíba, como requisito para a obtenção do grau de Bacharel em Engenharia de Computação

Orientador: Tiago Maritan Ugulino de Araújo

João Pessoa, PB  
Junho - 2018

**Catálogo na publicação**  
**Seção de Catalogação e Classificação**

S2371 Santos, Leandro Henrique de Souza.

LibrasTube: Um Aplicativo para a Criação de Janelas de Libras Sincronizadas com Vídeos do YouTube / Leandro Henrique de Souza Santos. - João Pessoa, 2018.  
41 f.

Orientação: Tiago Maritan Ugulino de Araújo.  
Monografia (Graduação) - UFPB/CI.

1. Libras, VLibras, Sincronização de Libras. I. Araújo, Tiago Maritan Ugulino de. II. Título.

UFPB/BC



CENTRO DE INFORMÁTICA  
UNIVERSIDADE FEDERAL DA PARAÍBA

Trabalho de Conclusão de Curso de Engenharia de Computação intitulado ***Libras-Tube: Um Aplicativo para a Criação de Janelas de Libras Sincronizadas com Vídeos do YouTube*** de autoria de Leandro Henrique de Souza Santos, aprovada pela banca examinadora constituída pelos seguintes professores:

---

Prof. Dr. Tiago Maritan Ugulino de Araújo  
Universidade Federal da Paraíba

---

Prof. Dr. Thaís Gaudencio do Rêgo  
Universidade Federal da Paraíba

---

Prof. Dr. Lincoln David Nery e Silva  
Universidade Federal da Paraíba

João Pessoa, 17 de Junho de 2018

*"Inclusão é o processo de trazer para o nosso mundo pessoas que, de alguma maneira, foram retiradas dele." (Nara Marcelino)*



À minha mãe por ser a pessoa mais trabalhadora que conheço.

## AGRADECIMENTOS

À professora Thaís Gaudencio do Rêgo, que nem é gente, é anjo. Pelo esforço extraordinário para transformar o Centro de Informática da UFPB em um lugar melhor e pelo apoio incessante aos alunos da instituição.

Ao professor Eudisley Gomes dos Anjos, por ser um exemplo de conduta e dedicação que sempre me serviu como inspiração.

Ao professor Tiago Maritan Ugolino de Araújo, por aceitar me orientar e pelo conhecimento transferido nesse importante projeto para minha formação acadêmica.

A todos os bons professores do centro de informática da UFPB que mantiveram seus papéis de ensinar e motivar os alunos de forma ética e justa.

Aos meus colegas que passaram madrugadas fazendo trabalhos e estudando para chegarmos onde chegamos.

## RESUMO

As Tecnologias da Informação e da Comunicação, quando são desenvolvidas, geralmente não endereçam as necessidades das pessoas com deficiência. Em razão disso, é necessário adaptar ou desenvolver novas tecnologias para reduzir as barreiras de acesso. Este trabalho endereça o problema de inclusão e acesso à informação das pessoas surdas em plataformas multimídia como o YouTube. Através de um aplicativo para Android chamado LibrasTube, este trabalho propõe uma arquitetura de tradução automática e sincronização de janelas de Libras através de legendas criadas por ferramentas de reconhecimento de fala. Assim, usuários surdos poderão assistir vídeos do YouTube de forma mais inclusiva através da adição de janelas de Libras nos vídeos. Na realização deste trabalho, foi verificado que a sincronização dessas janelas com o vídeo original é algo alcançável. Porém, o desenvolvimento de tecnologias de resumos textuais é necessário para que essa solução possa, de fato, atuar com uma qualidade satisfatória.

**Palavras-chave:** Libras, reconhecimento automático de fala, sincronização de janela de Libras, YouTube API, VLibras API.

## ABSTRACT

Information and Communication Technologies, when developed, generally do not address the needs of people with disabilities. As a result, new technologies need to be adapted or developed to reduce access barriers for people with disabilities. This paper addresses the problem of inclusion and access to information for deaf people on multimedia platforms such as YouTube. Through an Android application called LibrasTube, this work proposes an architecture of automatic translation and synchronization of Libras' windows through subtitles created by speech recognition tools. Therefore, deaf users will be able to watch YouTube videos more inclusively by adding Libras' windows to videos. In the realization of this work, it was verified that the synchronization of these windows with the original video is something reachable, however, the development of technologies for textual summaries is necessary so that this solution can, in fact, work with a satisfactory quality.

**Key-words:** Libras, automatic speech recognition, Libras synchronization, YouTube API, VLibras API

## LISTA DE FIGURAS

1	Exemplo de Ocorrência de Regionalismo em Libras (PEREIRA, 2010) . . .	20
2	Exemplo de Tradução de Português para Glosa em Libras . . . . .	21
3	Exemplo de tradução para Libras em tempo real [TV Justiça] . . . . .	22
4	Diagrama de blocos de um sistema de reconhecimento contínuo de fala (VEIGA, 2013) . . . . .	23
5	Arquitetura VLibras (FALCÃO, 2014) . . . . .	26
6	Arquitetura proposta . . . . .	28
7	Tela inicial do LibrasTube . . . . .	30
8	Fluxo principal do LibrasTube . . . . .	31
9	Captura de tela da janela de Libras sincronizada com vídeo oriundo do YouTube . . . . .	32
10	Lista parcial de legendas disponíveis no YouTube para um vídeo . . . . .	33
11	Prioridade da mais alta até a mais baixa na escolha da legenda. . . . .	33
12	Ajustes da janela de Libras . . . . .	34

## LISTA DE TABELAS

1	Exemplo de plataformas viáveis para a criação de janelas de Libras . . . .	18
2	Comparação da legenda gerada pelo YouTube através de um vídeo com boa qualidade de áudio . . . . .	35
3	Comparação da legenda gerada pelo YouTube através de um vídeo com má qualidade de áudio . . . . .	36

## LISTA DE ABREVIATURAS

ASR – Reconhecimento Automático de Fala (“Automatic Speech Recognition”)

API – Interface de Programação de Aplicativos (“Application Programming Interface”)

eMAG – Modelo de Acessibilidade em Governo Eletrônico

IBGE – Instituto Brasileiro de Geografia e Estatística

LIBRAS – Língua Brasileira de Sinais

LGP – Língua Gestual Portuguesa

NLP – Processamento de Linguagem Natural (“Natural-Language Processing”)

OMS – Organização Mundial da Saúde

PNS – Pesquisa Nacional de Saúde

TIC - Tecnologia da Informação e Comunicação

URL – Localizador Padrão de Recursos (“Uniform Resource Locator”)

WCAG – Diretrizes de Acessibilidade ao Conteúdo Web (“Web Content Accessibility Guidelines”)

## Conteúdo

<b>1</b>	<b>INTRODUÇÃO</b>	<b>16</b>
1.1	Identificação do Problema . . . . .	16
1.2	Objetivo geral . . . . .	17
1.3	Objetivos específicos . . . . .	18
1.4	Estrutura da monografia . . . . .	18
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b>	<b>20</b>
2.1	Libras . . . . .	20
2.2	Tradução simultânea de Libras no Brasil . . . . .	21
2.3	Reconhecimento Automático de Fala . . . . .	22
<b>3</b>	<b>TRABALHOS RELACIONADOS</b>	<b>24</b>
3.1	Ferramentas de tradução automática para Libras . . . . .	24
3.2	Suíte VLibras . . . . .	25
3.3	VLibras-Vídeo . . . . .	26
<b>4</b>	<b>SOLUÇÃO PROPOSTA</b>	<b>28</b>
4.1	Aplicativo LibrasTube . . . . .	29
<b>5</b>	<b>APRESENTAÇÃO E ANÁLISE DOS RESULTADOS</b>	<b>32</b>
5.1	Resultados da geração automática de legendas no YouTube . . . . .	32
5.2	Resultados da geração da janela de Libras através da suíte VLibras . . . . .	35
<b>6</b>	<b>CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS</b>	<b>38</b>
	<b>REFERÊNCIAS</b>	<b>39</b>

# 1 INTRODUÇÃO

## 1.1 Identificação do Problema

Segundo o último censo do Instituto Brasileiro de Comunicação e Estatística (IBGE), realizado em 2010, cerca de 5,1% da população brasileira possui algum tipo de deficiência auditiva [1]. O número de pessoas com deficiência auditiva severa, caracterizado por ter grande dificuldade ou não conseguir de modo algum ouvir, é de cerca de 1,12% da população. Dados mais recentes da Pesquisa Nacional de Saúde (PNS) de 2013 corroboram com os dados de 2010 do IBGE, estimando que 1,1% da população possui deficiência auditiva caracterizada como surdez nos dois ouvidos, surdez em um ouvido e audição reduzida no outro ouvido, ou ainda audição reduzida em ambos os ouvidos [10].

Em termos mundiais, a Organização Mundial da Saúde (OMS) estima que, em 2018, mais de 5% da população, o equivalente a 466 milhões de pessoas, possui deficiência auditiva caracterizada por perda auditiva superior a 40 decibéis (dB) na orelha de maior capacidade auditiva em adultos, ou perda auditiva maior que 30 dB em crianças [4]. Além disso, a OMS estima que, em 2050, esse número irá aumentar para 900 milhões de pessoas.

Diante desses dados e com o crescimento do número de pessoas que acessam à Internet no Brasil e no mundo, milhões de pessoas com deficiência auditiva podem acabar excluídas desse processo, devido a existência de poucos recursos de acessibilidade que permitam que elas acessem informações na Internet. Para reduzir esse problema, algumas iniciativas do Governo brasileiro foram desenvolvidas com o objetivo de promover acessibilidade para pessoas com deficiência. Um exemplo é a Lei 10.098 [7] de dezembro de 2000 que estabeleceu normas gerais e critérios básicos para a promoção da acessibilidade das pessoas com deficiência. Além disso, também foi criado o Modelo de Acessibilidade em Governo Eletrônico (eMAG) [5] que consiste em um conjunto de recomendações para padronizar e facilitar a implementação do processo de acessibilidade em sites e portais do governo brasileiro.

Com isso, algumas tecnologias vêm sendo desenvolvidas para tornar a informação mais acessível para as pessoas surdas, como, por exemplo, os tradutores automáticos de Português para Língua Brasileira de Sinais (Libras). Alguns exemplos desses tradutores automáticos para Libras são o VLibras [15], HandTalk [6] e ProDeaf [11]. Esses tradutores são utilizados geralmente para traduzir automaticamente o texto de diversos sites em português para animações em Libras. O VLibras, por exemplo, é atualmente utilizado nos sites do Governo Brasileiro, da Câmara dos Deputados, do Senado Federal, entre outros, enquanto que o Prodeaf e Handtalk são utilizados geralmente em sites comerciais. No entanto, não foram encontradas ferramentas que disponibilizem conteúdos em línguas de sinais em serviços de *streaming* de vídeo como, por exemplo, Youtube e Netflix. Esses

serviços estão em crescente expansão e a plataforma mais utilizada, Youtube, divulgou, em 2017, que possuía 1,5 bilhão de usuários ativos mensalmente e os mesmos assistiam, em média, mais de uma hora de conteúdo na plataforma apenas em dispositivos móveis [22] [14].

Um dos recursos de acessibilidade providos pelo YouTube é um serviço de adição de legendas ocultas geradas a partir de softwares de reconhecimento automático de fala. A transcrição descritiva de áudio gravado faz parte da recomendação 5.2 do eMAG, em conformidade com o Guia de Acessibilidade de Conteúdo Web (WCAG) [5]. No entanto, o Capítulo VII da Lei 10.098, que trata da acessibilidade nos sistemas de comunicação e sinalização, determina, no Art. 19, que: “Os serviços de radiodifusão sonora e de sons e imagens adotarão plano de medidas técnicas com o objetivo de permitir o uso da linguagem de sinais ou outra subtítuloção, para garantir o direito de acesso à informação às pessoas portadoras de deficiência auditiva, na forma e no prazo previstos em regulamento” [7]. Assim, o YouTube carece de uma ferramenta que disponibilize esses conteúdos em línguas de sinais, para que consiga promover uma melhor acessibilidade para pessoas com deficiência auditiva.

O uso de legendas para prover acessibilidade para pessoas com deficiência auditiva geralmente não é suficiente, uma vez que mesmo após muitos anos de escolaridade, as pessoas com deficiência auditiva geralmente têm dificuldade em compreender textos na língua oral de seu país [27]. Dessa forma, a melhor forma de assegurar que textos se tornem significativos a surdos é interpretá-los na língua de sinais [30].

Desta forma, o uso da Língua Brasileira de Sinais para tornar o conteúdo de plataformas de *streaming* de vídeo mais acessíveis, é um requisito necessário. Para endereçar esse problema, e como consequência, minimizar os problemas de acesso à informação das pessoas surdas em plataformas de *streaming* de vídeo, a proposta deste trabalho é investigar os principais problemas relacionados a geração de conteúdos em Libras nas plataformas de *streaming* de vídeo, bem como desenvolver um protótipo funcional para amenizar os problemas encontrados. Uma breve lista de plataformas que poderiam usar nossa proposta está descrita e caracterizada na Tabela 1

## 1.2 Objetivo geral

O objetivo geral deste trabalho é analisar os principais desafios acerca da implementação de soluções para tratar a acessibilidade em plataformas de *streaming* de vídeo utilizando janelas de Libras, de forma a tornar conteúdos audiovisuais mais acessíveis para pessoas surdas. Além disso, também pretende-se desenvolver um aplicativo que permita a tradução de vídeos do YouTube para Libras e sua apresentação de forma sincronizada com o vídeo.

**Tabela 1: Exemplo de plataformas viáveis para a criação de janelas de Libras**

<b>Plataformas</b>	<b>Tipos de Vídeo</b>	<b>Características</b>
<b>YouTube</b>	Variados	- Suporte à adição de arquivos de legenda - Mais de 1 bilhão de closed captions geradas
<b>Netflix, HBO, Hulu</b>	Filmes e séries	- Variado catálogo com legendas adicionadas em português
<b>Udacity, Udemy, Coursera</b>	Vídeo-aulas	- Geralmente só há um falante, o que facilita a conversão de fala para texto

### **1.3 Objetivos específicos**

1. Implementar um aplicativo Android capaz de pesquisar e selecionar vídeos do YouTube para a tradução;
2. Obter a transcrição do vídeo selecionado e convertê-la para uma sequência de glosas em Libras;
3. Gerar e exibir os sinais de Libras em uma janela mixada ao vídeo original selecionado;
4. Sincronizar a janela de Libras com o vídeo do YouTube.

### **1.4 Estrutura da monografia**

Esta monografia está organizada em cinco capítulos estruturados da seguinte forma:

- Capítulo 1: abordou o problema de pesquisa a ser tratado no escopo desta monografia, além de descrever os objetivos gerais e específicos;

- Capítulo 2: apresenta uma fundamentação teórica sobre os principais conceitos relacionados ao trabalho como a Língua Brasileira de Sinais, a tradução simultânea de Libras no Brasil e o reconhecimento automático de voz.
- Capítulo 3: apresenta os trabalhos relacionados ao trabalho aqui desenvolvido. Será discutido aplicações de tradução de português para Libras e, de forma mais detalhes, a suíte VLibras e o VLibras-Vídeo.
- Capítulo 4: explana a solução proposta, LibrasTube, descrevendo-a em suas principais funcionalidades.
- Capítulo 5: apresenta e discute os principais resultados obtidos com o desenvolvimento da solução.
- Capítulo 6: apresenta as considerações finais e algumas propostas de trabalhos futuros.

## 2 FUNDAMENTAÇÃO TEÓRICA

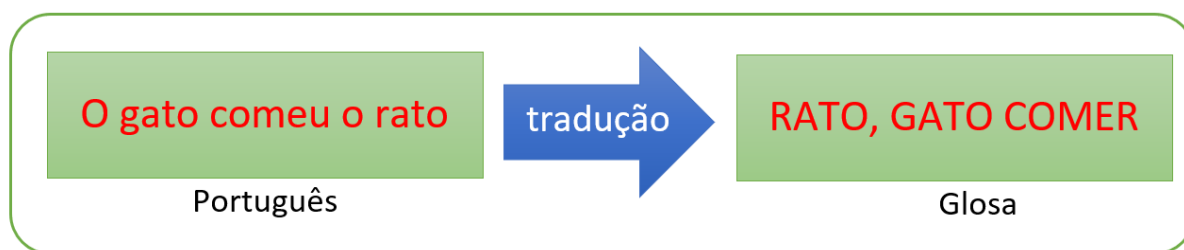
### 2.1 Libras

Libras é a segunda língua oficial do Brasil de acordo com a Lei 10.436 de 2002, consistindo de um sistema linguístico de natureza visual-motora oriunda de comunidades de pessoas surdas no Brasil para transmissão de ideias e fatos [8]. Os sinais de Libras são formados a partir de combinação das formas e movimentos das mãos e expressões faciais [30]. Além disso, a Língua Brasileira de Sinais não é universal em todos os estados brasileiros. Os sinais de Libras acabam sofrendo influências de suas regiões, assim como acontece com a língua Portuguesa [30]. Esse regionalismo da Libras adiciona uma variedade de sinais que torna a tradução de textos em sinais um desafio ainda maior. Um exemplo desse regionalismo pode ser visualizado na Figura 1, onde a palavra "verde" é sinalizada de maneira distinta entre as cidades do Rio de Janeiro, São Paulo e Curitiba.



Figura 1: Exemplo de Ocorrência de Regionalismo em Libras (PEREIRA, 2010)

Além disso, a Libras não é uma simples gestualização da língua portuguesa, mas sim uma língua à parte e, por isso, ela se difere da Língua Gestual Portuguesa (LGP) utilizada para a comunicação entre pessoas surdas em Portugal. Dessa forma, não basta apenas saber os sinais para utilizá-la, pois conhecer sua estrutura gramatical própria é fundamental para o correto entendimento pelos usuários da língua. A exemplo disso e como é demonstrado na Figura 2, o texto original em Português difere-se da representação textual em Libras, denominada glosa.



**Figura 2: Exemplo de Tradução de Português para Glosa em Libras**

Há diversos fatores que diferenciam a forma e estrutura gramatical entre a língua portuguesa e a Libras. Entre esses fatores, pode-se destacar:

- Os verbos aparecem sempre no infinitivo;
- Os pronomes pessoais não são representados, sendo necessário apontar a quem se refere;
- Os pronomes pessoais, possessivos e demonstrativos não possuem indicações de gênero;
- Os nomes próprios não necessariamente possuem sinais próprios. Nesse caso, faz-se a datilologia do sinal (isto é, sinaliza-se o nome letra a letra).

## 2.2 Tradução simultânea de Libras no Brasil

Atualmente, intérpretes da Língua Brasileira de Sinais atuam na tradução simultânea em atividades oficiais do Senado Federal, TV Senado, TV Justiça e Institutos Legislativos Brasileiro. Em programas de televisão, a imagem do intérprete é comumente sobreposta ao vídeo original no canto inferior direito, como pode ser visto na Figura 3 que apresenta a utilização de um intérprete de Libras na TV Justiça.

Uma característica importante do processo de tradução simultânea é que o intérprete precisa manter a informação em Libras o mais próximo possível da informação original. No entanto, pessoas falam, em média, de duas a três palavras por segundo em português [21]. Um único sinal de Libras, por sua vez, tem duração média de dois a três segundos, o que torna mais difícil a tradução simultânea de forma sincronizada. Assim, um dos focos da interpretação deve ser o significado da mensagem e não as palavras a serem traduzidas, levando em consideração que o tempo é crítico nessas situações [28]. Dessa forma, os sistemas de tradução automática de textos para sinais em Libras devem ser capazes de entender essas características intrínsecas a língua de sinais para tentar contornar esses problemas. Portanto, afim de sincronizar sinais de Libras com a fala, poderá ser necessário o uso resumos e simplificações textuais.



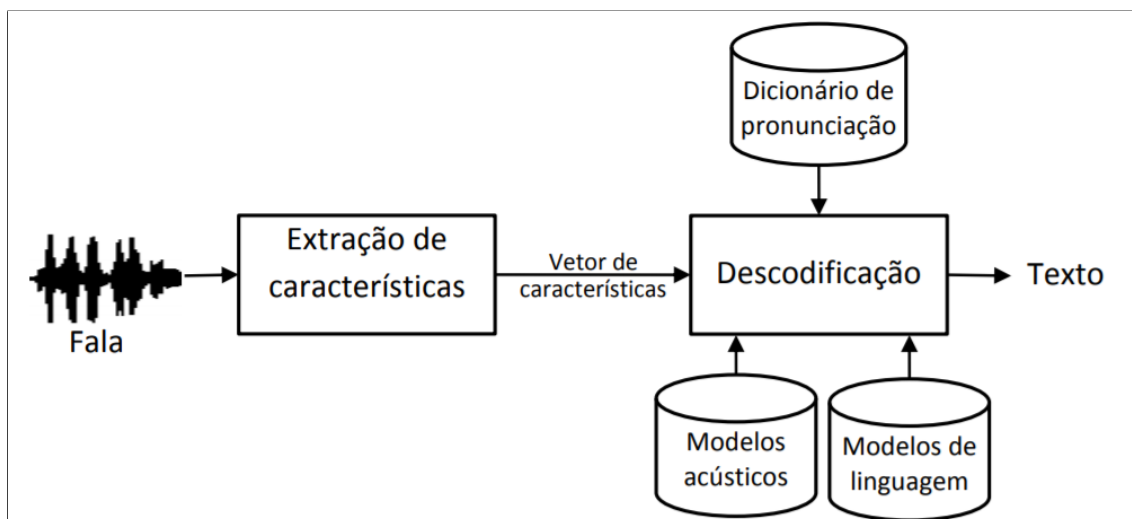
**Figura 3: Exemplo de tradução para Libras em tempo real [TV Justiça]**

### **2.3 Reconhecimento Automático de Fala**

O reconhecimento automático de fala é o processo de usar o software para converter as palavras faladas de uma pessoa em uma transcrição de texto. Esses softwares conseguem, através da captura de um sinal acústico representativo da fala, determinar as palavras que foram faladas pela correspondência de padrões [19]. Os sistemas de reconhecimento automático de fala geralmente possuem um conjunto de modelos acústicos e de idiomas armazenados em banco de dados. Esses modelos são o resultado de treinamento e de regras de interpretação do idioma que são comparados aos sinais capturados como pode ser visto Figura 4.

De acordo com a Figura 4, o módulo de análise de extração de características tem como proposta a parametrização da fala em uma sequência de vetores, chamados de vetores de características, que contém informações relevantes sobre a expressão. Essas características devem estar dentro do domínio espectral contido na voz e devem prover uma boa distinção em modelos estatísticos [29]. Essas características, então, serão comparadas a modelos acústicos de palavras e modelos de linguagem para realizar a conversão das características da fala para sentenças que irão compor o texto final.

Um dos principais fatores de complicação no reconhecimento da fala contínua é que os limites das palavras são muitas vezes difíceis de serem classificados [19]. Por exemplo,



**Figura 4: Diagrama de blocos de um sistema de reconhecimento contínuo de fala (VEIGA, 2013)**

softwares de reconhecimento de fala têm dificuldade de diferenciar o limite de uma palavra de um sub-palavra (por exemplo, um fonema). Dessa forma, técnicas de processamento de linguagem natural (PLN) são utilizadas para otimizar o processo de conversão.

Baseando-se nessas técnicas, softwares de reconhecimento automático de fala têm conseguido alcançar bons resultados em vídeos de diferentes perfis. Entre os sistemas de reconhecimento automático de fala mais utilizadas está a Google Cloud Speech-to-Text que é capaz de transcrever áudio em mais de 110 idiomas [2]. Além disso, em 2017, o YouTube divulgou que já possui mais de 1 bilhão de legendas ocultas geradas automaticamente a partir de falas [9].

Esta seção abordou conceitos importantes para a fundamentação deste trabalho. Entre os pontos a serem destacados, estão: as características da língua brasileira de sinais e algumas das distinções da língua portuguesa, a utilização de janelas de Libras com intérpretes reais na TV Brasileira, e a fundamentação do processo de reconhecimento de fala. Na próxima seção, serão abordados alguns trabalhos relacionados que são correlatos ao trabalho aqui desenvolvido.

## 3 TRABALHOS RELACIONADOS

### 3.1 Ferramentas de tradução automática para Libras

Existem diversas ferramentas com o objetivo de realizar a tradução de português para Libras. A maioria dessas ferramentas fazem a tradução apenas a partir de textos, algumas contam ainda com a funcionalidade de traduzir informações textuais de websites para sinais de Libras. Outras ferramentas, no entanto, são capazes de realizar a tradução para Libras também através de áudio. De forma geral, as ferramentas geram animações que são capazes de executar os movimentos dos sinais de Libras.

Entre as principais diferenças dessas ferramentas, estão a capacidade de realizar um passo intermediário de tradução do conteúdo da mensagem para Glosa (representação textual de sinais) ao invés de gerar os sinais diretamente a partir de estrutura gramatical portuguesa. Além disso, algumas ferramentas geram um avatar 3D que pode ser rotacionado para visualizar os sinais em diferentes ângulos.

Atualmente, na literatura, existe uma vasta variedade de ferramentas de tradução automática para Libras, segue alguns exemplos com uma breve descrição de suas características:

- **F-LIBRAS:** ambiente virtual cujas principais funções são a gravação, editoração e visualização de gestos e movimentos tridimensionais da língua de sinais através de técnicas de realidade virtual [24];
- **FALIBRAS:** um tradutor de português (escrito ou falado) para língua brasileira de sinais em formato gestual e animado com integração ao navegador web Firefox. O tradutor utiliza-se de técnicas de aprendizado de máquina e possui o potencial de traduzir qualquer língua escrita ou falada para qualquer língua de sinal [20];
- **HandTalk:** empresa fundada em 2012 com foco em tradução de sites para Libras. Seu tradutor é capaz de gerar sinais a partir de texto e áudio e conta com aplicativos para dispositivos móveis [6];
- **POLI-LIBRAS:** um sistema de código aberto capaz de fazer a tradução Português-Libras através de uma animação 3D. O diferencial deste projeto é a sua modularização, o que facilita a integração com outros projetos com finalidades semelhantes [26];
- **ProDeaf:** um conjunto de softwares capazes de traduzir texto e voz de português para Libras. O ProDeaf possui versões para web e dispositivos móveis, além de ser capaz de traduzir websites [11];

- **Projeto RyBená:** um projeto de acessibilidade para surdos que iniciou-se com uma solução que converte automaticamente mensagens de texto para sinais de Libras no celular. O RyBená TV foi desenvolvido a partir do projeto inicial com o objetivo de traduzir legendas ocultas em TV Digital, porém a tradução é feita em português sinalizado, não em Glosa, e não há sincronização com as marcações de texto das legendas ocultas [12][18].

### 3.2 Suíte VLibras

Dentre todos esses sistemas de tradução automáticos supracitados, o VLibras é o único capaz de receber arquivos de legenda para a geração de Libras [23]. Além disso, ele é um software livre que possui um dicionário de sinais que está em contínua expansão, e a capacidade de realizar um passo intermediário de tradução *text-to-gloss* (isto é, uma tradução do texto para glosa em Libras). Em razão disso, ele foi a plataforma escolhida para apoiar o desenvolvimento deste trabalho e será analisado de uma forma um pouco mais aprofundada.

A Suíte VLibras é um conjunto de ferramentas computacionais que faz a tradução automática de conteúdos digitais como, por exemplo, textos, áudios e vídeos para Libras, tornando esses conteúdos acessíveis para pessoas surdas. Além da capacidade de traduzir para diferentes tipos de mídias, o VLibras também possui suporte para diferentes plataformas como, por exemplo, os sistemas operacionais Windows, Linux, Android e iOS, e nos navegadores Google Chrome, Firefox e Safari.

Como pode ser visto na arquitetura da suíte VLibras apresentada na Figura 5, o VLibras gera a tradução das animações em Libras a partir de uma sequência de passos que se inicia com a obtenção do conteúdo digital a ser traduzido. Caso o conteúdo não seja obtido diretamente através de texto, um módulo de extração irá extrair o texto a partir de um vídeo, áudio ou legenda. A tradução para glosa ocorre com o auxílio de um conjunto de regras de tradução que são obtidas através de uma ferramenta colaborativa chamada WikiLIBRAS, que é responsável por definir regras de tradução e sinais animados de Libras. Após a realização da tradução, a animação poderá ser criada e, se preciso, sincronizada com os tempos de fala do seu conteúdo de origem.

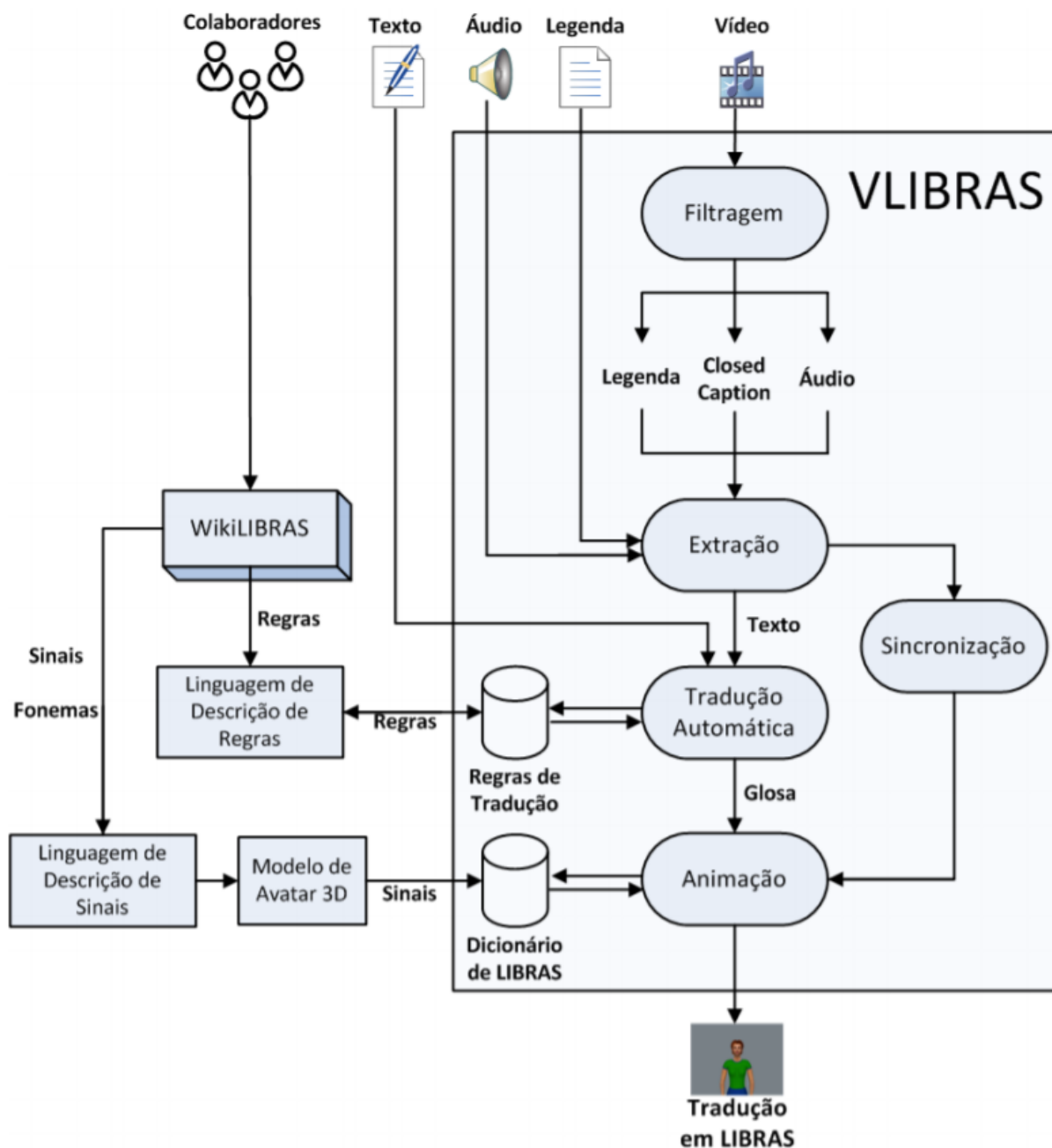


Figura 5: Arquitetura VLibras (FALCÃO, 2014)

### 3.3 VLibras-Vídeo

O VLibras-Vídeo é uma ferramenta contida na suíte VLibras capaz de gerar janelas de Libras a partir de legendas enviadas pelos usuários ou geradas pelo seu software interno de reconhecimento de fala. Esta ferramenta pode ser acessada através do seu *website* [16] ou através de sua API pública [17]. O principal diferencial do VLibras-Vídeo em relação a outras ferramentas de tradução para Libras é a funcionalidade de sincronizar a janela de Libras com as marcações de tempo das legendas.

A API do VLibras-Vídeo conta com alguns *endpoints* utilizados para acessar dife-

rentes funcionalidade da ferramenta. Entre as principais funcionalidades da API, pode-se destacar:

- **Vídeo-Legenda:** recebe um arquivo de vídeo e um arquivo de legenda. Após o término da tradução do texto e a geração dos sinais, um novo vídeo é retornado com uma janela de Libras incorporada ao vídeo original de forma sincronizado com os tempos da legenda;
- **Vídeo:** recebe apenas um arquivo de vídeo e faz o reconhecimento de fala para gerar legendas. Após a geração da legenda, o processo ocorre como declarado no item supracitado;
- **Legenda:** recebe apenas um arquivo de legenda que será usado para gerar uma janela de Libras sincronizada com os tempos da legenda. Um novo vídeo é retornado apenas com os sinais de Libras;

Nos testes realizados durante o desenvolvimento deste trabalho, apenas o primeiro *endpoint* está respondendo as requisições corretamente como referenciado em sua documentação [17]. Ainda assim, a ferramenta é a única com a funcionalidade de gerar janelas de Libras sincronizadas com arquivos de legendas. Esta funcionalidade pode ser utilizada em diversas aplicações como a aplicação proposta neste trabalho.

Esta seção descreveu alguns trabalhos desenvolvidos correlatos com o tema e solução proposta deste trabalho. A próxima seção irá descrever a solução desenvolvida e irá detalhar a aplicação criada, denominada LibrasTube, que foi desenvolvida com o objetivo de provar os conceitos da proposta apresentada.

## 4 SOLUÇÃO PROPOSTA

Muitas ferramentas de tradução automática de português para Libras tratam apenas a tradução de textos para língua de sinais, não dando a devida atenção à sincronização desses sinais com o texto ou áudio original. Dessa forma, legendas e textos com marcações de tempo acabam perdendo sua principal função que é a de sincronizar a língua, seja escrita ou sinalizada, com a fala.

Nossa solução propõe o uso de legendas para a criação de janelas de Libras que serão exibidas de forma conjunta e sincronizada com o vídeo original. A aquisição da legenda pode ser realizada através da adição direta de um arquivo de legenda ou através do uso de ferramentas de conversão de fala para textos com marcações de tempo. Dessa forma, será possível gerar um novo vídeo a partir do vídeo original com a adição de uma janela de Libras sincronizada com as falas do vídeo original.

Como pode ser visto na Figura 6, caso o vídeo possua uma legenda, o módulo de tradução automática para língua de sinais pode gerar uma janela de Libras e adicioná-lo ao vídeo original de forma direta. Caso o vídeo não possua um arquivo de legenda próprio, o sistema poderá extrair as falas com seus respectivos tempos de duração e iniciação a partir de softwares de reconhecimento automático de fala (ASR). Com a obtenção da legenda e a conversão posterior para janela de Libras sincronizada, o usuário poderá assistir o vídeo original e entender informações sonoras a partir da janela de Libras adicionada.

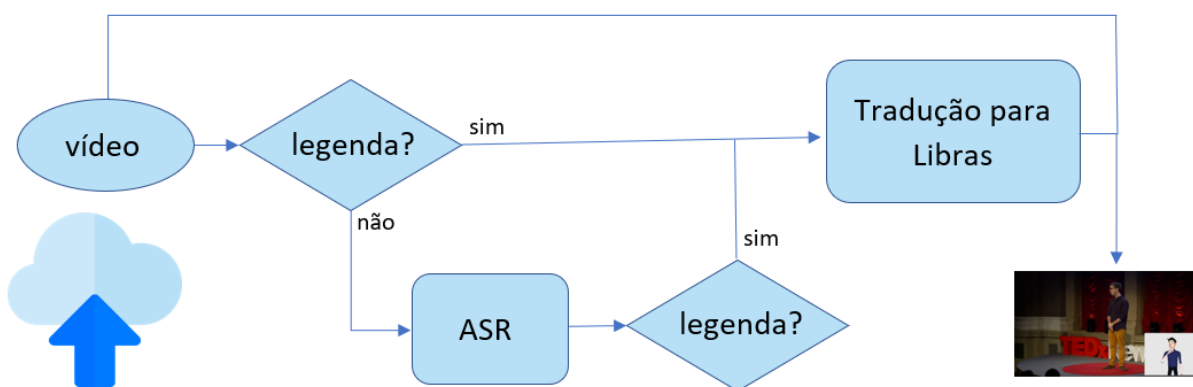


Figura 6: Arquitetura proposta

Dessa forma, espera-se que vídeos oriundos de diferentes plataformas possam ser utilizados para a criação de janelas de Libras sincronizadas com o vídeo original, gerando maior acessibilidade a pessoas surdas.

## 4.1 Aplicativo LibrasTube

Com o objetivo de construir uma prova de conceito da proposta, um aplicativo, denominado LibrasTube, foi desenvolvido para a plataforma *mobile* Android. A proposta do LibrasTube é traduzir automaticamente vídeos do YouTube para Libras, mantendo o sincronismo com o vídeo. O aplicativo faz uso de três itens: o vídeo a ser traduzido, a API do YouTube que dá acesso ao arquivo de legenda (seja enviado pelo usuário ou autogerado) e a API do VLibras para fazer a geração das janelas de Libras já sincronizadas com a legenda. Dessa forma, é possível recriar vídeos do YouTube com janelas de Libras adicionadas e sincronizadas com o áudio original.

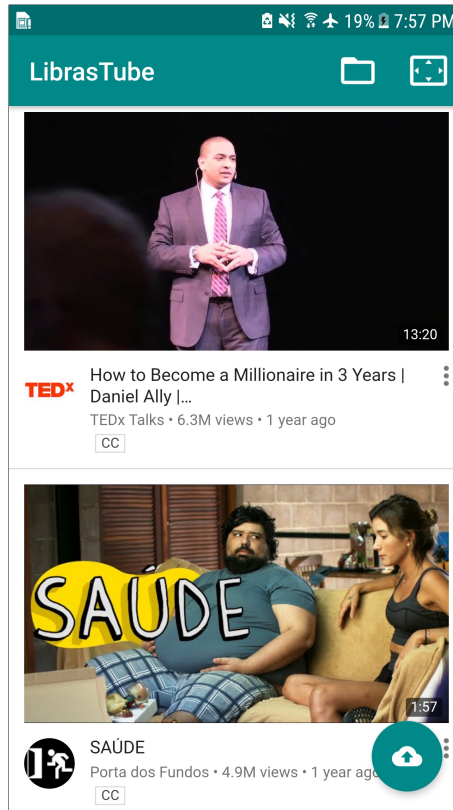
Como pode ser visto na Figura 7, o aplicativo exibe um componente do Android chamado *WebView* que é capaz de apresentar o conteúdo de um site (neste caso, o YouTube) diretamente dentro de um aplicativo. Dessa forma, o usuário pode navegar pelo site do YouTube dentro do aplicativo LibrasTube e escolher o vídeo que deseja realizar a tradução.

Como discutido na Seção 2.3 deste trabalho, o YouTube, através de seu software de reconhecimento automático de fala, é capaz de gerar legendas em mais de 110 idiomas. Além disso, a plataforma possui mais de 1 bilhão de legendas geradas automaticamente através dessa tecnologia. Assim, o LibrasTube pode utilizar as legendas geradas automaticamente pelo YouTube assim como arquivos de legenda adicionados na plataforma por usuários para a geração de janelas de Libras.

O principal fluxo de execução do aplicativo pode ser visualizado na Figura 8. A execução dos eventos realizados nas três telas é descrito da seguinte forma:

1. Com o vídeo a ser convertido selecionado, o usuário poderá clicar no botão flutuante na parte inferior direita da tela para dar início ao processo de *download* e envio do vídeo e da legenda extraída para fazer a tradução;
2. Após a primeira etapa ser realizada, o usuário poderá visualizar os status dos vídeos selecionados através do botão “Vídeos” (representado por um ícone que remete a um diretório) localizado na parte superior direita da tela;
3. Na tela “Vídeos”, o usuário poderá visualizar todos os vídeos que ele selecionou assim como seus status. É nesta tela que o usuário poderá selecionar o vídeo a ser assistido já com a janela de Libras sincronizada.

É importante ressaltar que os vídeos e legendas extraídos do YouTube precisam ser baixados para serem enviados ao servidor do VLibras devido a restrições da própria API do VLibras. No entanto, esses arquivos são excluídos do dispositivo do usuário



**Figura 7: Tela inicial do LibrasTube**

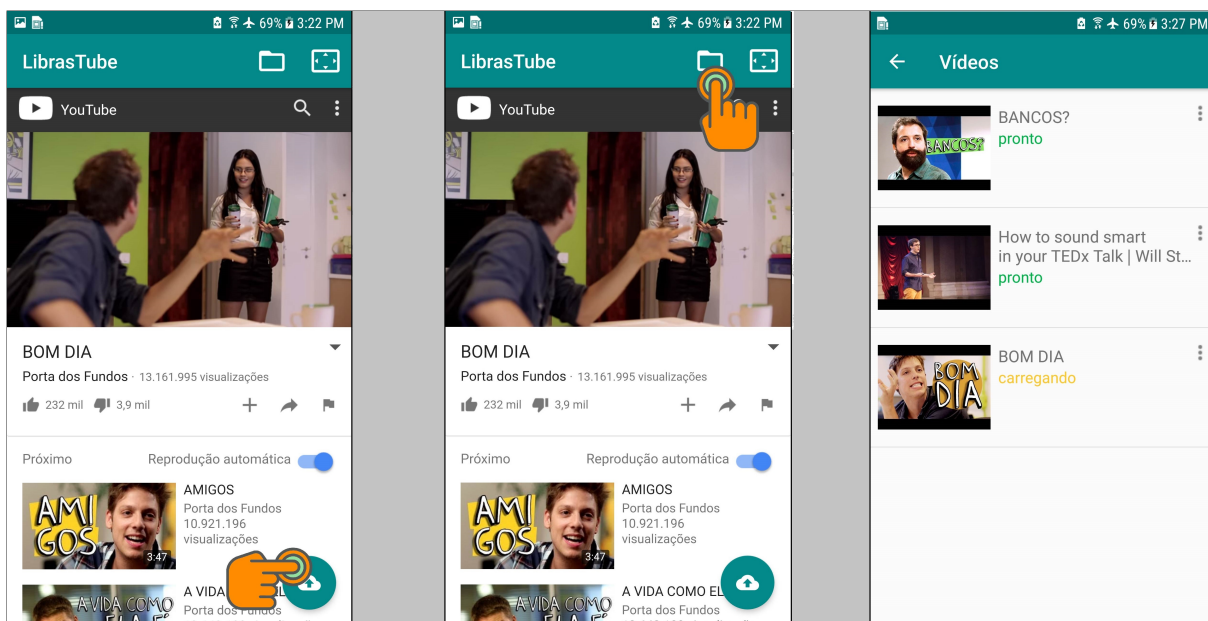
imediatamente após o envio, de forma que o usuário não possa ter acesso a eles para outros fins.

Após a execução do fluxo principal do LibrasTube demonstrado na Figura 8, o usuário pode assistir o vídeo através de *streaming* realizado a partir dos servidores do VLibras. Para isso, basta clicar em qualquer vídeo com o status concluído e o vídeo será executado no dispositivo do usuário como demonstrado na Figura 9.

O vídeo demonstrado na Figura 9 possui áudio e legenda original em Inglês. No entanto, a API do YouTube permite a tradução da legenda para diversos idiomas caso seja necessário. Além disso, o YouTube possui uma lista de possíveis aquisições de legendas como legendas adicionadas pelo usuário e legendas autogeradas a partir de softwares de reconhecimento de fala.

A Figura 10 mostra um exemplo de uma lista resumida de legendas disponíveis no YouTube para determinado vídeo. Neste caso, existem duas legendas adicionadas pelo usuário do YouTube que estão marcadas pelo tag “track”. As demais legendas são traduções das legendas originais. A lista pode, ainda, possuir uma legenda com o valor “kind=asr” que representa que a legenda foi gerada a partir de um software de reconhecimento automático de fala.

Devido à variedade de opções de legendas disponíveis no YouTube, o aplicativo



**Figura 8: Fluxo principal do LibrasTube**

procura selecionar a legenda de maior qualidade. A seleção do tipo legenda para a criação da janela de Libras obedece a ordem de prioridade que pode ser visualizada na Figura 11. Nota-se que legendas adicionadas pelo usuário em outros idiomas que podem ser traduzidas para o português possuem maior prioridade que as legendas autogeradas a partir do áudio original em português. A lista de prioridades foi criada a partir de análises empíricas de variados vídeos.

Para adequação a preferências do usuário, a exibição da janela de Libras pode ser personalizada antes da seleção do vídeo através do menu de ajustes. Como pode ser visto na Figura 12, o usuário pode escolher, através do botão de ajuste na parte superior direita da tela, a posição da janela de Libras assim como o tamanho que ela irá ser exibida.

Para avaliar a solução proposta, no Capítulo 5 será apresentado e discutido os principais resultados encontrados. Os resultados discutidos estão relacionados a qualidade e viabilidade da sincronização com o áudio original assim como a qualidade de legendas geradas por ferramentas de reconhecimento de voz.



**Figura 9:** Captura de tela da janela de Libras sincronizada com vídeo oriundo do YouTube

## **5 APRESENTAÇÃO E ANÁLISE DOS RESULTADOS**

O aplicativo LibrasTube foi capaz de integrar tecnologias existentes como ferramentas de reconhecimento automático de fala assim como ferramentas de tradução de texto em português para Libras. As ferramentas utilizadas no desenvolvimento do LibrasTube foram, respectivamente, Cloud Speech-to-Text API da Google e VLibras API. Essas ferramentas foram avaliadas em trabalhos anteriores, e encontram-se em funcionamento há alguns anos. Assim, não é o objetivo deste trabalho avaliá-las individualmente e de forma extensiva, mas sim, verificar seus desempenhos no cenário de criação e sincronização de janelas de Libras.

### **5.1 Resultados da geração automática de legendas no YouTube**

Como explicado na Seção 4, o YouTube é capaz de exibir legendas a partir de arquivos adicionados por usuários assim como arquivos gerados a partir de softwares de reconhecimento automático de fala. Dessa forma, será brevemente discutido o segundo caso onde a legenda fica a cargo do software de reconhecimento de fala do YouTube.

Diferentemente de como é feito em traduções simultâneas realizada por pessoas, softwares de tradução de texto para Libras não conseguem corrigir falhas no texto original como frases mal formadas e ocultação de palavras. Dessa forma, toda falha encontrada em uma legenda gerada a partir do reconhecimento de fala será propagada para a janela de Libras.

```

<transcript_list docid="2216951700531275620">
<format fmt_code="1" default="true"/>
<format fmt_code="vtt"/>
<format fmt_code="ttml"/>
<track id="0" name="" lang_code="en" lang_original="English" lang_translated="Inglês" cantran="true" vss_id=".en"/>
<track id="1" name="" lang_code="pt-BR" lang_original="Português (Brasil)" lang_translated="Português (Brasil)" lang_default="true" cantran="true" vss_id=".pt-BR"/>
<target id="84" urlfrag="&lang=af" lang_code="af" lang_original="Afrikaans" lang_translated="Africâner" vss_id=".af"/>
<target id="43" urlfrag="&lang=sq" lang_code="sq" lang_original="Shqip" lang_translated="Albanês" vss_id=".sq"/>
<target id="5" urlfrag="&lang=de" lang_code="de" lang_original="Deutsch" lang_translated="Alemão" vss_id=".de"/>
<target id="67" urlfrag="&lang=am" lang_code="am" lang_original="አማርኛ" lang_translated="Amárico" vss_id=".am"/>
<target id="50" urlfrag="&lang=ar" lang_code="ar" lang_original="العربية" lang_translated="Árabe" vss_id=".ar"/>
<target id="81" urlfrag="&lang=hy" lang_code="hy" lang_original="Հայերեն" lang_translated="Armênio" vss_id=".hy"/>
<target id="68" urlfrag="&lang=az" lang_code="az" lang_original="Azərbaycan" lang_translated="Azerbaijano" vss_id=".az"/>
<target id="53" urlfrag="&lang=eu" lang_code="eu" lang_original="Euskara" lang_translated="Basco" vss_id=".eu"/>
<target id="35" urlfrag="&lang=bn" lang_code="bn" lang_original="বাংলা" lang_translated="Bengali" vss_id=".bn"/>
<target id="45" urlfrag="&lang=be" lang_code="be" lang_original="Беларуская" lang_translated="Bielorrusso" vss_id=".be"/>
<target id="86" urlfrag="&lang=my" lang_code="my" lang_original="မြန်မာ" lang_translated="Birmanês" vss_id=".my"/>
<target id="71" urlfrag="&lang=bs" lang_code="bs" lang_original="Босански" lang_translated="Bósnio" vss_id=".bs"/>
<target id="25" urlfrag="&lang=bg" lang_code="bg" lang_original="Български" lang_translated="Búlgaro" vss_id=".bg"/>
<target id="54" urlfrag="&lang=kn" lang_code="kn" lang_original="ಕನ್ನಡ" lang_translated="Canarim" vss_id=".kn"/>
<target id="51" urlfrag="&lang=ca" lang_code="ca" lang_original="Català" lang_translated="Catalão" vss_id=".ca"/>
<target id="89" urlfrag="&lang=kk" lang_code="kk" lang_original="Қазақ Тілі" lang_translated="Cazaque" vss_id=".kk"/>
<target id="12" urlfrag="&lang=pt" lang_code="pt" lang_original="Português" lang_translated="Português" lang_default="true" vss_id=".pt"/>
<target id="76" urlfrag="&lang=ky" lang_code="ky" lang_original="Кыргызча" lang_translated="Quirquiz" vss_id=".ky"/>
</transcript_list>

```

Figura 10: Lista parcial de legendas disponíveis no YouTube para um vídeo

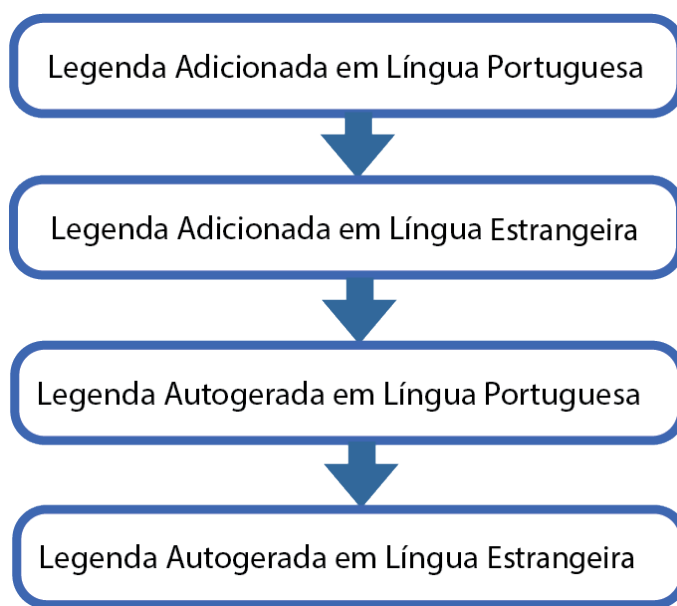
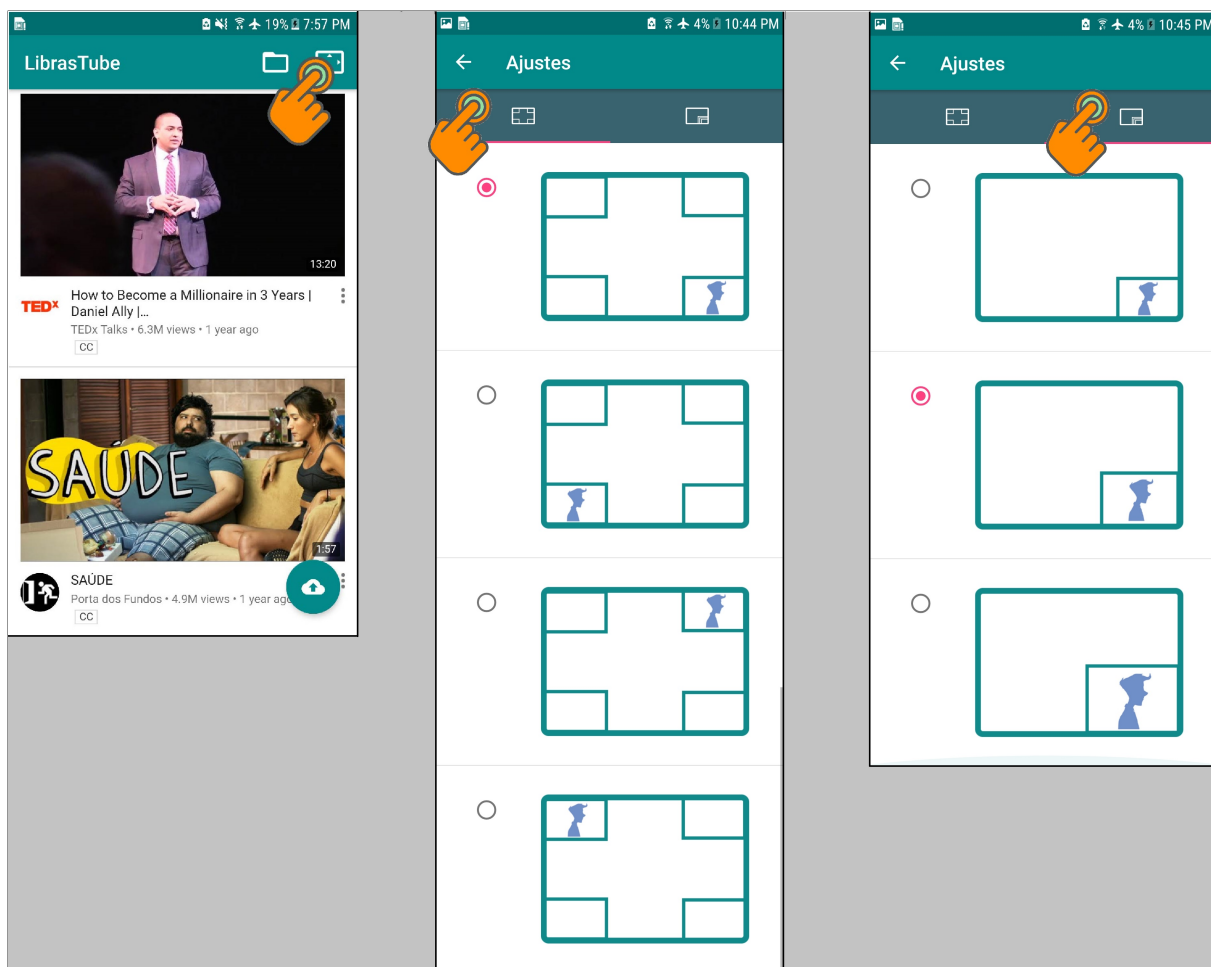


Figura 11: Prioridade da mais alta até a mais baixa na escolha da legenda.

A qualidade das legendas geradas por esses softwares depende diretamente da qualidade do áudio além de outros fatores como a qualidade da articulação e pronúncia das palavras e a clareza na formação de frases pelos falantes. Consequentemente, a qualidade da janela de Libras exibida pelo aplicativo LibrasTube pode variar bastante de acordo com o vídeo selecionado para a tradução, podendo, até mesmo, ser incompreensível.

Em casos de legendas geradas a partir de vídeos com boa qualidade de áudio, a expectativa é que existam poucos erros. Como pode ser visto na Tabela 2, um trecho de uma reportagem do telejornal intitulado Jornal Nacional foi utilizado para comparar a legenda gerada pelo YouTube com a legenda esperada. Nesta situação específica, o software de reconhecimento de fala teve uma taxa de similaridade com o texto original de 95.5%, de acordo com o algoritmo de similaridade da ferramenta CopyLeaks [3]. Assim, houve falhas apenas no reconhecimento de um nome próprio, “Marielle”, e na separação das letras “e” e “a”. Além disso, o software não foi capaz de aplicar regras de pontuação



**Figura 12: Ajustes da janela de Libras**

de forma correta. No entanto, a quantidade de falhas não prejudicou o texto a ponto de torná-lo incompreensível.

Em casos de legendas geradas a partir de vídeos com má qualidade de áudio, espera-se, no entanto, que haja uma quantidade de erros maior. Como pode ser visto na Tabela 3, um trecho de uma aula de Libras no YouTube foi utilizado para comparar a geração de legendas. Nessa situação, o vídeo selecionado possui baixa qualidade de áudio e o falante não fala com tanta clareza e qualidade de pronúncia de um apresentador de telejornal. Assim, é verificado que há uma taxa de similaridade de apenas 46.9%, o que dificulta o entendimento e se caracteriza como uma legenda imprópria para a criação de uma janela de Libras.

Dessa forma, existe a expectativa que o aplicativo LibrasTube tenha uma baixa qualidade na criação de janelas de Libras nas situações em que o áudio original se apresenta com uma baixa qualidade. Por outro lado, em situações com boa qualidade de áudio, é esperado a criação de janelas de Libras com um bom nível de entendimento pelos usuários. No entanto, será necessária uma avaliação com pessoas surdas para verificar se a quantidade de erros, mesmo que pequena, possa inviabilizar a compreensão das

**Tabela 2: Comparação da legenda gerada pelo YouTube através de um vídeo com boa qualidade de áudio**

Legenda Esperada	Legenda Gerada
<p>Uma multidão se reuniu no centro para pedir a investigação e a punição dos autores do crime.</p>	<p>Uma multidão se reuniu no centro para pedir a investigação <b>ea</b> punição dos autores do crime.</p>
<p>Em Belo Horizonte, manifestantes levaram faixas e cartazes para protestar contra a violência sofrida por jovens negros e fazer uma homenagem a vereadora Marielle Franco.</p>	<p>Em belo horizonte manifestantes levaram faixas e cartazes para protestar contra a violência sofrida por jovens negros e fazer uma homenagem a vereadora <b>mary l</b> franco.</p>
<p>As reações foram além das fronteiras do país. Em Portugal, as pessoas reuniram numa vigília no centro de Lisboa para homenagear a vereadora.</p>	<p>As reações foram além das fronteiras do país em portugal as pessoas reuniram numa vigília no centro de lisboa para homenagear a vereadora.</p>
<p>A ONG Redes da Maré deixou uma homenagem à mulher que levou à voz das minorias para muito além da favela onde ela nasceu.</p>	<p>A ong redes da maré deixou uma homenagem à mulher que levou à voz das minorias para muito além da favela onde ela nasceu.</p>

informações. A verificação e qualificação da capacidade de entendimento de pessoas surdas será proposto em trabalhos futuros. Esta análise da geração de janelas de Libras em situações que haja falhas nas animações oriundas do texto original é de grande importância haja visto que essas ferramentas estão frequentemente sujeitas a este tipo de erro.

## 5.2 Resultados da geração da janela de Libras através da suíte VLibras

A suíte VLibras foi selecionada neste trabalho por ser o único software capaz de receber arquivos de legenda para a geração de Libras sincronizadas com os tempos das falas. Apesar de atender os objetivos gerais deste trabalho, alguns problemas foram encontrados durante o seu uso e serão aqui discutidos.

De acordo com a documentação da API VLibras, uma das funcionalidades presentes é a de geração de um vídeo de Libras a partir, apenas, de um arquivo de legenda. No entanto, esta funcionalidade não está funcionando como documentada até o momento da publicação deste trabalho. Assim, torna-se necessário o envio do arquivo de vídeo juntamente com o arquivo de legenda. Em consequência disso, o processo de geração

**Tabela 3: Comparação da legenda gerada pelo YouTube através de um vídeo com má qualidade de áudio**

Legenda Esperada	Legenda Gerada
<p>Vou apresentar para vocês o alfabeto e os números em datilologia: o alfabeto com as mãos, o alfabeto manual.</p> <p>Daí você então tem primeiro o sinal do alfabeto manual. Então, quando você quiser que alguém fale uma palavra em datilologia você coloca o “a”, “b”, “c” e treme os dedos assim. Dessa forma, você está dizendo: faça em datilologia.</p> <p>Geralmente quando não tem um sinal para uma palavra ou então uma pessoa ou nome de alguém, aí você coloca. O meu por exemplo: “l”, “u”, “i”, “z”.</p> <p>Acontece que para você memorizar, decorar, compreender todo o alfabeto e os números, têm sido muito cansativo, têm sido às vezes difícil para os alunos, até pra mim.</p>	<p>Vou apresentar para vocês o alfabeto e os números em <b>da filologia</b> o alfabeto <b>as</b> mãos o alfabeto manual.</p> <p>Daí você então tem <b>primeiro</b> sinal do alfabeto manual então quando você quiser que alguém <b>fala</b> uma palavra em <b>tecnologia</b> você <b>colocou</b> a, b c e <b>três dias</b> assim dessa forma você está dizendo faça em <b>tecnologia</b>.</p> <p>Geralmente, quando não tem um sinal pra <b>falar então</b> uma pessoa <b>o</b> nome de alguém aí você coloca o meu por exemplo <b>ego e c</b>.</p> <p>Acontece <b>você analisar e</b> compreender todo o alfabeto <b>os</b> números têm sido muito cansativo tem sido às vezes difícil para os alunos até pra mim.</p>

da janela de Libras tornava-se mais lento, em virtude da necessidade de envio do vídeo. Além disso, a necessidade de baixar o arquivo de vídeo do YouTube para o dispositivo do usuário pode gerar problemas de direitos autorais.

De forma análoga, a documentação ainda define que o vídeo original poderia ser enviado através de sua URL. Assim, a API se encarregaria de fazer o *download* do vídeo original para a eventual geração de um novo vídeo com sua janela de Libras embutida. No entanto, essa funcionalidade não está presente, tornando-se necessário fazer o *upload* do vídeo original que pode demorar vários minutos e consumir pacote de dados do usuário, além de possivelmente comprometer a experiência do usuário.

Além disso, muitos dos sinais apresentados na janela de Libras são executados em uma velocidade mais lenta que outros sistemas de tradução de texto para Libras, o que pode acabar comprometendo a sincronização da janela de Libras com as falas, uma vez que a sinalização em Libras é geralmente mais lenta que as falas em línguas orais. Dessa

forma, isso pode resultar na perda de sincronização do vídeo com a janela de Libras.

Pode-se concluir, então, que melhorias nessas ferramentas de tradução de português para Libras são necessárias e cruciais para aplicações como o LibrasTube. Em especial, a melhoria da qualidade de sincronização de forma a encontrar o limiar que define a velocidade máxima de execução da animação sem perder sua inteligibilidade. Além disso, torna-se necessária a expansão de possibilidades de utilização desse tipo API, a exemplo, a possibilidade de enviar legendas sem a necessidade de enviar seus vídeos originais assim como a aceitação de legendas em diferentes formatos.

## 6 CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS

O estudo desenvolvido neste trabalho apresentou uma solução de acessibilidade para surdos a partir de vídeos que pode ser utilizada em diversas plataformas multimídia, especialmente as plataformas de *streaming* de vídeos. Para construir uma prova de conceito da proposta, foi desenvolvido um aplicativo para Android chamado LibrasTube.

O LibrasTube foi capaz de integrar legendas adicionadas ou geradas pelo YouTube com o sistema VLibras de tradução de texto para Libras. Assim, usuários surdos poderão assistir vídeos do YouTube e acompanhar o conteúdo sonoro através de uma janela de Libras adicionada no canto do vídeo.

No entanto, alguns desafios precisam ser resolvidos para que esta solução esteja em um nível aceitável de qualidade para ser utilizado por pessoas surdas. Dentre estes desafios, pode-se destacar:

- **Ferramenta de resumo textual:** um dos principais desafios na representação de sinais de Libras em sincronia com legendas é a sua adequação aos tempos de duração. A exibição de sinais em Libras costuma ser mais lenta que as palavras faladas, o que dificulta muito a sincronização dos sinais às falas. Assim, torna-se necessário ferramentas capazes de resumir textos de legendas de forma a permitir a adequação temporal desses sinais aos tempos das falas.
- **Maior dicionário de Libras:** apesar de softwares de tradução já possuírem dicionários com cerca de 15 mil sinais definidos, a língua portuguesa possui mais de 300 mil palavras. Dessa forma, é necessária a criação de mais sinais, assim como mapeamento de sinônimos e adição de gírias e expressões regionais.
- **Melhor qualidade de animação dos sinais:** devido a necessidade de acelerar as animações para se ajustar aos tempos de fala, alguns sinais podem ficar tão rápidos a ponto de se tornarem incompreensíveis. Assim, torna-se necessário o aperfeiçoamento desses sinais para que o usuário possa compreender o sinal mesmo que executado de forma mais acelerada.

Apesar deste trabalho ter conseguido abordar o tema e alcançar a realização da maior parte dos objetivos estabelecidos, há alguns trabalhos futuros importantes que devem ser realizados a fim de refinar a proposta apresentada. Entre os trabalhos futuros necessários, pode-se destacar:

- **Realizar testes de usabilidade, experiência de usuário e adaptação a legendas com falhas:** é importante que o software seja avaliado por usuários surdos

para a identificação de falhas de usabilidade. Caso seja necessário, a página web do YouTube poderia ser recriada através de API para melhor se adequar a usuários surdos. Além disso, é necessário verificar como usuários surdos se adaptam e compreendem sinais gerados a partir de legendas que foram autogeradas com erros, o que frequentemente acontece com legendas criadas a partir de ferramentas de reconhecimento de fala.

- **Criação de um sistema semiautomático de resumo de legendas:** o uso de um sistema capaz de resumir legendas de forma semiautomática poderia melhorar a qualidade de sincronização de janelas de Libras de forma considerável. Este é um trabalho crucial para o sucesso de soluções com janelas de Libras sincronizadas como o LibrasTube.
- **Expansão para outras plataformas:** o YouTube é a plataforma de vídeo mais utilizada atualmente, o que maximiza o número de usuários beneficiados com a adição da janela de Libras. Contudo, há diversas outras plataformas como Netflix, Udemy e Udacity que poderiam, notavelmente, se beneficiar de uma solução semelhante e com uma qualidade possivelmente superior devido a facilidade de controle de qualidade dos seus vídeos e legendas.

## REFERÊNCIAS

- [1] **CENSO DEMOGRÁFICO 2010. Características gerais da população, religião e pessoas com deficiência.** Disponível em: <ftp://ftp.ibge.gov.br/Censos/Censo Demografico 2010/ Caracteristicas Gerais Religiao Deficiencia/ caracteristicas religiao deficiencia.pdf>. Acesso em: 15 mar 2018.
- [2] **CLOUD SPEECH API.** Disponível em: <https://cloud.google.com/speech/> . Acesso em: 16 abr 2018.
- [3] **CopyLeaks.** Disponível em: <https://copyleaks.com/compare>, Acesso em: 16 jun 2018.
- [4] **Deafness and hearing loss.** Disponível em: <http://www.who.int/mediacentre/factsheets/fs300/en/>, Acesso em: 15 mar 2018.
- [5] **eMAG - Modelo de Acessibilidade em Governo Eletrônico.** Disponível em: <https://www.governoeletronico.gov.br/eixos-de-atuacao/governo/acessibilidade/emag-modelo-de-acessibilidade-em-governo-eletronico>, Acesso em: 15 mar 2018.
- [6] **Hand Talk.** Disponível em: <https://www.handtalk.me/sobre>, Acesso em: 6 jun 2018.
- [7] **Lei Número 10.098 de dezembro de 2000.** Disponível em: <http://www.planalto.gov.br/ccivil\_03/leis/L10098.html>, Acesso em: 15 mar 2018.
- [8] **Lei Número 10.436 de abril de 2002.** Disponível em: <http://www.planalto.gov.br/ccivil\_03/Leis/2002/L10436.html>, . Acesso em: 16 mar 2018.
- [9] **One billion captioned videos.** Disponível em: <https://youtube.googleblog.com/2017/02/one-billion-captioned-videos.html>, Acesso em: 15 abr 2018.
- [10] **Percepção do estado de saúde, estilos de vida e doenças crônicas.** Disponível em: <ftp://ftp.ibge.gov.br/PNS/2013/pns2013.pdf>, Acesso em: 15 mar 2018.
- [11] **ProDeaf.** Disponível em: <http://www.prodeaf.net/pt-br/OQueE>, Acesso em: 6 jun 2018.
- [12] **RyBena.** Disponível em: <http://portal.rybena.com.br/site-rybena/>, Acesso em: 6 jun 2018.

- [13] **TREINO NÃO SUPERVISIONADO DE MODELOS ACÚSTICOS PARA RECONHECIMENTO DE FALA.** Disponível em: <[https://estudogeral.sib.uc.pt/bitstream/10316/24262/1/Modelos básicos para reconhecimento de fala.pdf](https://estudogeral.sib.uc.pt/bitstream/10316/24262/1/Modelos_basicos_para_reconhecimento_de_fala.pdf)>, Acesso em: 05 jun 2018.
- [14] **Updates from VidCon: more users, more products, more shows and much more.** Disponível em: <<https://youtube.googleblog.com/2017/06/updates-from-vidcon-more-users-more.html>>, Acesso em: 15 mar 2018.
- [15] **VLibras.** Disponível em: <<http://www.vlibras.gov.br/>>, Acesso em: 05 jun 2018.
- [16] **VLibras-Vídeo.** Disponível em: <<http://video.vlibras.gov.br/>>, Acesso em: 7 jun 2018.
- [17] **VLibras-Vídeo API.** Disponível em: <<https://www.softwarepublico.gov.br/gitlab/vlibras/vlibras-video-container/tree/master>>, Acesso em: 7 jun 2018.
- [18] Lóscio B. F. AMORIM M. L. C., Assad R. **RyBenáTV - Solução para Acessibilidade de Surdos para TVDigital.** 2010.
- [19] Chigie B. **ASR - Automatic speech recognition,** 1997.
- [20] CORADINE L. C. BRITO P. H. S., FRANCO N. **FALIBRAS: uma Ferramenta Flexível para Promover Acessibilidade de Pessoas Surdas.** 2012.
- [21] Vanessa. CELESTE Letícia Côrrea COSTA, Luanna Maria Oliveira. OLIVEIRA. **Metodologias de análise da velocidade de fala: um estudo piloto,** 2016.
- [22] WEST Darrell M. **The Evolution of Video Streaming and Digital Content Delivery.** Disponível em: <[https://www.brookings.edu/wp-content/uploads/2016/06/West Evolution-of-VideoStreaming-and-Digital-Content-Delivery Final.pdf](https://www.brookings.edu/wp-content/uploads/2016/06/West_Evolution-of-VideoStreaming-and-Digital-Content-Delivery_Final.pdf)>, Acesso em: 3 mar. 2018.
- [23] FALCÃO Eduardo de Lucena. **Deaf Accessibility as a Service: uma Arquitetura Escalável e Tolerante a Falhas para o Sistema de Tradução VLIBRAS,** 2005.
- [24] BAPTISTA F. **F-Libras - Ambiente Integrado de Ensino-Aprendizagem para Língua Brasileira de Sinais.** 2007.
- [25] PEREIRA Graciele Kerlen. **LIBRAS (Língua Brasileira de Sinais),** 2010.
- [26] KOGA M. L. JANUARIO J. F, LEITE L. A. F. **POLI-LIBRAS - Um Tradutor de Português para Libras.** 2010.

- [27] STUMPF Marianne Rossi. **Aprendizagem de Escrita de Línguas de Sinais pelo Sistema SignWriting: *Linguas de Sianis no Papel e no Computador***. PhD thesis, Universidade Federal do Rio Grande do Sul, 2005.
- [28] RONICE MÜLLER QUADROS. **O tradutor e intérprete de língua brasileira de sinais e língua portuguesa**, 2004.
- [29] FURUI Sadaoki. *Automatic Speech and Speaker Recognition*, chapter **Pattern Recognition Approach**. Kluwer Academic Publishers, 2012.
- [30] PEREIRA Simone Rodrigues. **Os Processos de Alfabetização e Letramento em Libras: Um Percorso Semiótico**, 2009.