

Desenvolvimento e avaliação de tecnologia para tradução de áudios para Libras em dispositivos móveis.

Trabalho de Conclusão do Curso Superior de Tecnologia em Sistemas Para Internet

Jaqueline Paz Bonoto

Orientador: Rodrigo Prestes Machado

¹Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul (IFRS)
Campus Porto Alegre
Av Cel Vicente, 281, Porto Alegre – RS – Brasil

jaqueline.pazbonoto@gmail.com, rodrigo.machado@poa.ifrs.edu.br

***Resumo.** O crescimento do uso de smartphones para comunicação e para acesso a conteúdos de diversas naturezas é uma realidade. Embora smartphones contemplem aspectos de acessibilidade e inclusão, ainda há o que melhorar no que se refere à interação de pessoas surdas e ouvintes por meio de mensagens de áudio, por exemplo. Pensando nessa problemática, foi realizado um estudo acerca da Suite VLibras, buscando desenvolver um app para que pessoas com deficiência auditiva e pessoas ouvintes possam interagir usando conteúdos de áudio, especialmente em redes sociais. Para atingir esse objetivo foi implementado um sistema utilizando tecnologias Android, IBM Watson e VLibras. Por fim foi avaliada a assertividade na conversão de conteúdos de áudios para Libras e foi observado índice de eficácia de 50%.*

1 - Introdução

No Brasil, aproximadamente 10 milhões de pessoas possuem deficiência auditiva; isso equivale a 5% da população (IBGE, 2010). Esses indivíduos interagem com o mundo principalmente por meio da visão e uma de suas mais importantes manifestações linguísticas é a Língua Brasileira de Sinais (Libras), sendo o português sua segunda língua (FURLAN, 2016). Considerando a necessidade de inclusão dessa comunidade, o conceito de acessibilidade definido na Constituição Federal Brasileira diz que sistemas de comunicação e de informação devem prover condição para sua utilização por parte de pessoas com deficiência (BRASIL, 2016).

Pesquisa sobre tipo de dispositivos utilizados para acesso à internet aponta que em uma amostragem de 100 pessoas, 31 usam computador e 98 usam *smartphones* para uso da *web* fora do âmbito profissional (CETIC, 2018). Surdos também utilizam telefones para se comunicar, mas existem limitações no que diz respeito à tradução para Libras. Aplicativos

para tradução em Libras, como HandTalk¹ e o VLibras-App², oferecem uma solução para traduzir textos e áudios inseridos manualmente pelos usuários em seus aplicativos; no entanto, não tornam acessíveis sons recebidos de outros recursos dos aparelhos, como os de redes sociais. Segundo Torres *et al.* (2018) aplicativos como Whatsapp aumentam a interação entre surdos e ouvintes devido aos recursos do aplicativo (mensagens, videochamadas, imagens); por outro lado, como há restrições quanto à tradução de áudios para Libras, as pessoas com deficiência auditiva acabam por interagir menos que aquelas que não a tem.

Dessa forma, o principal objetivo deste trabalho é construir um aplicativo capaz de capturar conteúdo em áudio de qualquer aplicativo rodando em *smartphones* Android e traduzir para Libras representado-o pela notação em glosa³. Para atingir o resultado final foram definidos e realizados os seguintes objetivos secundários: estudar trabalhos relacionados e tecnologias que poderiam contribuir para a solução do problema; desenvolver um aplicativo de tradução; validar a proposta de solução por meio de análise da eficácia da tradução do conteúdo de áudio para Libras (glosa).

2 - Referencial Teórico

Para modelar a solução pretendida, o referencial teórico é estruturado em dois momentos: investigação dos trabalhos relacionados e estudo de conceitos e tecnologias. Na primeira parte, são estudados projetos cujos propósitos estão alinhados à tradução de mídias de áudio para Libras em *smartphones* e na *web*, com o objetivo de compreender métodos, testes e resultados. Em seguida, na seção “Conceitos e tecnologias”, serão abordados recursos tecnológicos pertinentes ao desenvolvimento de aplicativos Android, transcrição com IBM Watson e consumo de API de tradução para Libras. O referencial teórico foi utilizado como base para o desenvolvimento do método de trabalho deste projeto.

2.1 - Trabalhos Relacionados

Esta seção do artigo foi elaborada a partir de uma revisão de literatura nas bases Scopus e Google Scholar buscando pelos termos “VLibras”, “libras smartphone” e “Hand Talk”. Por ser um conjunto de ferramentas públicas para tradução de português para Libras, a Suite VLibras essa foi escolhida como alvo do estudo. A premissa era encontrar projetos recentes que implementam VLibras e que tiveram como resultado valor tecnológico agregado a esse conjunto de ferramentas. Após essa pesquisa, foi possível identificar projetos que contribuíssem no alcance dos objetivos deste artigo.

Costa *et al.* 2019 buscaram transformar o “Libras *translation Rules*” - *machine translation* baseada em regras aplicada à Suite VLibras - em um repositório de regras de tradução. A ideia era possibilitar o acoplamento de outras *machines translation* de modo que os

¹ Disponível em <<https://www.handtalk.me/>>. Acesso em: 05 nov, 2019.

² Disponível em <https://play.google.com/store/apps/details?id=com.lavid.vlibrasdroid&hl=pt_BR>. Acesso em: 05 nov, 2019.

³ Glosa é uma notação usada para descrever sinais, provendo o significado do sinal usando palavras de um idioma, como o português (PAIVA, 2019).

conteúdos pudessem ser traduzidos de diversos idiomas falados para Libras. Para comprovar sua eficácia, foram realizados testes computacionais aplicando as métricas BLEU (*bilingual evaluation understudy*) e WER (*word error rate*) e testes com usuários para medir o desempenho do protótipo entregue.

Paiva (2019) contribuiu para a formalização de regras morfosintáticas para tradução automática de português para Libras e realizou a implementação de sistema que aplica essas regras. Os objetivos da sua tese orbitam em torno do fato de que Libras é uma língua visual e espacial, e que por isso não há unanimidade na comunidade científica no que diz respeito à representação escrita de seus sinais. As regras foram estudadas e implementadas em código Python e o processo de validação aplicou testes com usuários, selecionando conjuntos de frases alinhadas a essas regras. Por fim, a investigação se utilizou de métricas como SER (*sentence error rate*), BLEU e WER.

Em outro projeto, a proposta foi o desenvolvimento de um protótipo de aplicativo móvel para conversão de voz em texto e texto em voz. Para validar o modelo construído, Furlan (2016) divulgou o aplicativo em redes sociais: teve 81 *downloads* e captou 56 avaliações de usuários no Google Play, obtendo a média de 4,7 pontos em um escore de 1 a 5. Para seu desenvolvimento, foram empregados Android SDK, JavaScript e Angular.js, HTML e CSS, Ionic, APIs Apache Cordova e Web Speech.

Os trabalhos referenciados trouxeram aspectos fundamentais para o desenvolvimento deste projeto: Costa *et al.* 2019 elucidou o funcionamento da Suite VLibras e introduziu métricas para avaliação da tradução; Paiva (2019) contribuiu com conjuntos de frases com análise de regras morfológicas capazes de ser aplicados em testes de objetivos semelhantes; e Furlan (2016) trouxe um exemplo prático de criação de um protótipo de aplicativo para comunicação entre surdos e ouvintes. Embora haja farta literatura e muitos projetos relacionados à acessibilidade de surdos no uso de *smartphones*, não se destaca a produção de estudos com o propósito de traduzir conteúdos de áudios advindos de recursos de *smartphones*, objetivo central deste projeto.

2.2 - Conceitos e tecnologias

Uma vez que o objetivo geral deste trata-se de melhorar a inclusão entre surdos e ouvintes no uso de recursos em *smartphones* por meio do desenvolvimento de um app, nesta subseção aborda-se a Suite pública VLibras na busca pela compreensão dos recursos que ela disponibiliza para tradução de português para Libras. Optou-se desenvolver um aplicativo Android, para tanto este tópico também será abordado, e o IBM Watson entrou como alternativa para preparar o conteúdo a ser enviado para VLibras. Por fim, será descrita a seleção de conteúdos a serem traduzidos e, em seguida, estarão dispostas as métricas de assertividade que proporcionam as ferramentas para a verificação de resultados.

2.2.1 - VLibras

A Suite VLibras - também comumente chamada de VLibras - é um conjunto de ferramentas de código aberto desenvolvido na Universidade Federal da Paraíba (UFPB) que pretende

tornar dispositivos móveis e *websites* acessíveis para pessoas com deficiência auditiva por meio da tradução de conteúdos digitais de português para Libras. Neste projeto, foi utilizado o VLibras Service da Suíte VLibras:

De acordo com Costa *et al.* (2019), o VLibras Service oferece uma API (*application programming interface*) que pode ser utilizada gratuitamente. Ela recebe valores de entrada do tipo texto e pode entregar como valores de saída a glosa ou conteúdo multimídia acessível.

A API do VLibras Service⁴ oferece *endpoints* que podem ser consumidos externamente, possibilitando assim o uso de diversas linguagens para criação de uma aplicação funcional. Desse modo, a API VLibras caracteriza um ator externo ao sistema desenvolvido neste projeto, já que a aplicação irá enviar requisições *web* para esta API.

2.2.2 - Android

Nos últimos cinco anos, 24 milhões de brasileiros começaram a usar a internet por meio de dispositivos móveis que utilizam sistema operacional Android, segundo levantamento da Bain & Company⁵. Essa popularização foi determinante para sua inclusão neste projeto.

O sistema operacional Android viabiliza a utilização de recursos próprios dos dispositivos na implementação de projetos. De acordo com documentação disponível no *site* Android Developers⁶, permissões do tipo RECORD_AUDIO e STORAGE podem ser adicionadas ao app e implementadas por meio da classe MediaRecorder para utilizar recursos de gravação e armazenamento. Dessa maneira é possível gravar e armazenar mídias de áudio.

2.2.3 - IBM Watson

Watson é uma plataforma de serviços cognitivos da empresa IBM. Já aplicada em diversas áreas do setor público brasileiro, ela oferece recursos de aprendizado de máquina e ferramentas como o Speech to Text⁷.

Esse serviço possibilita a transcrição de conteúdos de áudio a partir da tecnologia de reconhecimento de fala. O Watson Speech to Text é capaz de realizar conversões em mais de 15 idiomas. Neste projeto, converteu-se somente o português. Uma vez que a API VLibras suporta apenas entrada de texto, pretendeu-se utilizar a Watson para transformar o áudio captado no *smartphone* em texto e, então, traduzi-lo para Libras.

⁴ Disponível em <<https://traducao2.vlibras.gov.br/docs>>. Acesso em: 31 out, 2020.

⁵ Disponível em <<https://www.bain.com/pt-br/insights/economic-and-social-impact-of-android-in-brazil/>> . Acesso em: 17 nov, 2020.

⁶ Disponível em <<https://developer.android.com/training/permissions/requesting>>. Acesso em: 31 out, 2019.

⁷ Disponível em <<https://www.ibm.com/br-pt/cloud/watson-speech-to-text>>. Acesso em: 14 out, 2020.

2.2.4 - Seleção de conteúdo para tradução para glosa.

Idealmente, cabe à solução proposta ser capaz de traduzir qualquer conteúdo recebido. No entanto, a fim de delimitar um conjunto finito de testes, foi necessário selecionar objetos de tradução. Considerando-se que a glosa é uma notação usada para descrever sinais da Libras, é possível dizer que “[...] estruturar glosas em posição adequada não é uma tarefa trivial” (PAIVA, 2019, p. 90). Diferentes pontos de atenção em relação à língua devem ser considerados, como: a necessidade de se manterem no infinitivo os verbos em Libras (glosa); a ausência de certos sinais em Libras para algumas palavras; a união, na língua de sinais, de quantificadores como “muito” ao substantivo; entre outros.

Paiva (2019) selecionou um conjunto de frases que contempla os aspectos mencionados. O Quadro 1 apresenta esse conjunto.

Quadro 1. Frases selecionadas para teste de tradução para Libras

Número da frase	Frase em português
1	O bebê é muito calmo à noite.
2	Muitas motos foram roubadas na cidade.
3	Existem muitos animais raros no zoológico.
4	Cuidado com o bebê e os filhotes.
5	Naquela rua tem muitas casas de madeira.
6	Na loja tem muitos produtos de limpeza.
7	O Papa Francisco é muito simpático.
8	Será que o vento sopra sempre na mesma direção?
9	O vento pode soprar em várias direções.
10	Maria comeu muitas maçãs.

Fonte: Adaptado de Paiva (2019).

2.2.5 - Métricas para índice de eficácia de tradução

Para validar a solução, utilizaram-se métricas para a avaliação de eficácia de tradução, posto que qualidade é um aspecto importante para o conteúdo da mensagem ser compreendido após ser submetido ao fluxo do aplicativo. Nesta seção, portanto, são abordadas métricas para este tipo de análise: a WER e a SER.

2.2.5.1 - WER (*word error rate*)

De acordo com Paiva (2019), *word error rate*, ou “taxa de erro de palavras” é uma métrica que considera o número de inserções, exclusões e substituições ao comparar duas frases. Para a obtenção do índice WER, usa-se a seguinte fórmula:

$$\text{WER} = \frac{\text{Substituições} + \text{inserções} + \text{exclusões}}{\text{Número de palavras de referência}}$$

Essa métrica pode ser aplicada automaticamente através de um algoritmo e tem como limitação a possibilidade de a tradução de um idioma para outro pode demandar alteração da posição das palavras, o que impacta diretamente no índice. De qualquer maneira, se utilizada para verificar a taxa de erro na transcrição de áudios, ela se mostra eficaz e responde de forma adequada no processamento da correspondência entre a transcrição e o conteúdo.

2.2.5.2 - SER (*sentence error rate*)

Sentence error rate (SER), em português “taxa de erro de sentenças”, é uma métrica para avaliar a eficácia de tradução de conteúdo. O índice SER representa o número de frases incorretas dividido pelo número total de frases testadas, como apresentada na fórmula a seguir.

$$\text{SER} = \frac{\text{Número de frases incorretas}}{\text{Número total de frases}}$$

Em sua metodologia de validação, Paiva (2019) aplica a métrica SER solicitando aos usuários que realizassem a categorização das frases do seguinte modo:

- 0 - Frase incompreensível.
- 1 - Frase compreensível, mas com problemas.
- 2 - Frase compreensível.

Caso a glosa oferecida pelo sistema não faça sentido de acordo com a análise do voluntário, ele categoriza a tradução como “0”. Caso o sentido da frase tenha sido mantido, porém não aparenta ser a tradução mais adequada, a categorização será de “1”. Se a frase é compreensível, é categorizada como “2”. A vantagem dessa métrica é que ela é facilmente calculada sem auxílio de algoritmos. A pessoa que avalia se a frase está correta leva em consideração a frase original e o resultado da tradução, indicando assim se o sentido da mensagem foi mantido.

3 - Método de Trabalho

Nesta seção é apresentada uma visão geral sobre a proposta de solução, bem como os passos para sua implementação e validação. Na etapa de implementação, foi desenvolvido o aplicativo conforme as tecnologias estudadas e selecionadas.

Já na etapa de validação, o intuito foi compreender o impacto do app considerando-se o problema que se propôs a resolver. Para isso, o resultado das traduções foram avaliados por pessoas bilíngues (fluentes em Libras e em português) e os resultados da transcrição de texto passaram pela avaliação de um algoritmo.

3.1 - Visão geral da solução

Na Figura 1, as ilustrações de tela exemplificam os passos que são realizados para traduzir um conteúdo de áudio para glosa conforme a proposta de solução. Cada tela indica um passo a ser executado, e os botões a seguir mencionados estão destacados por quadrados vermelhos: (1) o usuário inicializa a gravação de áudio clicando no botão Iniciar; (2) abre o aplicativo que contém o áudio a ser transcrito e o reproduz; (3) após a reprodução, retorna para o app de tradução e finaliza a gravação clicando no botão Parar; (4) solicita a tradução para glosa clicando no botão Traduzir para glosa; (5) visualiza em tela o conteúdo do áudio traduzido para Libras (glosa), podendo assim compreender a informação antes inacessível.



Figura 1. Protótipo para simulação de passos que o usuário irá seguir para tradução dos conteúdos de áudio para glosa.

3.1.1 - Implementação

Como o objetivo principal deste projeto é tornar acessíveis conteúdos de áudios por meio da criação de um aplicativo, a etapa de implementação é determinante na entrega da proposta de solução. O diagrama da Figura 2 demonstra como as solicitações do usuário acionam diferentes funcionalidades implementadas no código: o escopo da fase de implementação tem como base as três funcionalidades numeradas e destacadas em retângulos vermelhos. Elas estão organizadas da seguinte maneira: gravação de áudios do *smartphone* (1); transcrição de áudio para texto (2) e tradução do texto em português para Libras (3).

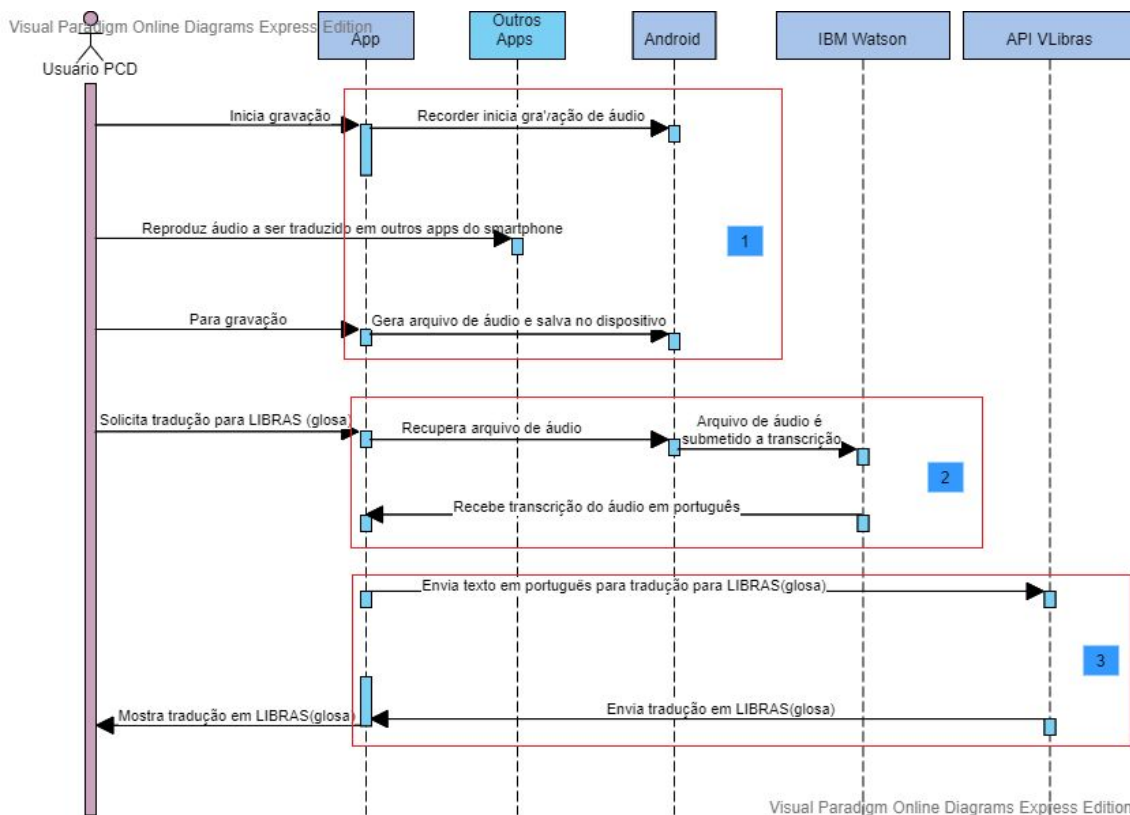


Figura 2. Diagrama com ações de usuário e ações do app.

Para o desenvolvimento da funcionalidade de gravação de áudio, foram utilizados recursos próprios do sistema operacional Android. Ao gravar e armazenar conteúdos de áudio foi necessário solicitar permissões do tipo `RECORD_AUDIO` e `STORAGE`. Após a implementação dessas permissões, foram adicionados os botões Iniciar gravação e Parar gravação. Desenvolvendo a lógica com a classe `MediaRecorder`, o botão Iniciar gravação aciona o microfone do *smartphone*, e o botão “Parar gravação” finaliza a gravação e salva o áudio na memória do *smartphone*, conforme demonstra a funcionalidade 1 da Figura 2.

O sistema Android é capaz de salvar diversos formatos de arquivos de áudio. No caso deste estudo, foi necessário aplicar uma extensão compatível com a plataforma IBM, que trataria o arquivo posteriormente. O formato escolhido, portanto, foi o `ogg`. A utilização desse recurso serviu aos propósitos do projeto e atendeu aos seguintes critérios: o áudio gravado pode ser encontrado no aparelho; o áudio pode ser reproduzido; a gravação capta sons emitidos por outros aplicativos; os conteúdos de áudio são claros e compreensíveis.

No desenvolvimento da funcionalidade 2 da Figura 2, preparou-se o conteúdo a ser consumido pela API VLibras, como a API possui *endpoints* para entradas de texto, e não de áudio, fez-se necessário transcrever o arquivo de áudio. Para isso foi utilizada a ferramenta Speech to Text da plataforma IBM Watson. Dessa forma, quando o usuário pressiona o botão Traduzir para glossa, o arquivo de áudio é enviado para o ator externo, que, por sua vez, devolve a transcrição do conteúdo por JSON (*JavaScript object notation*). Depois disso, o conteúdo já está apto para ser traduzido para glossa.

Por fim, conectou-se o sistema aos *endpoints* de tradução do VLibras. Assim, a transcrição recebida da IBM Watson é enviada a um *endpoint* da API para que os recursos retornem em glosa. O conteúdo em glosa é, então, exibido na tela, e o usuário pode acessá-lo em Libras, como mostra a funcionalidade 3 da Figura 2.

Uma vez implementada a solução proposta e desenvolvidas as funcionalidades necessárias para satisfazer aos critérios básicos estabelecidos, foi possível iniciar a etapa de validação conforme o que havia sido proposto.

3.2 - Validação

Para realizar a validação da solução, foram desenvolvidos dois tipos de testes. O primeiro, que avaliava a qualidade da tradução, foi realizado com pessoas bilíngues. O outro, responsável por verificar a eficácia da etapa de transcrição, foi realizado com o auxílio de um algoritmo WER. O objetivo da validação com usuários era aferir a qualidade da tradução por meio da utilização do recurso desenvolvido; já a análise da transcrição poderia responder se essa etapa impactou a qualidade da tradução. Para realizar a validação, foram considerados os seguintes aspectos: análise da tradução por pessoas bilíngues; análise da transcrição do áudio e aplicação de métricas.

3.2.1 - Testes com pessoas voluntárias

Na etapa de testes com usuários, as frases foram enviadas no formato de mensagem de áudio pelo aplicativo WhatsApp utilizando a voz da autora deste projeto. Cada uma delas foi submetida ao fluxo completo do app 3 vezes, variando o volume de áudio do *smartphone* (primeiro em 10%, depois em 40% e, por último em 70% da capacidade do dispositivo). Nessa alternância de volume de reprodução, percebeu-se que o volume em 40% apresentava menor ruído, o que proporcionava uma transcrição mais próxima ao conteúdo original. Os testes, portanto, foram realizados com volume de áudio de 40% e com a voz de uma única pessoa, a fim de reduzir o impacto da transcrição na análise geral da solução, bem como o escopo da validação. Os áudios tiveram duração de 3 a 4 segundos.

Realizaram a validação duas pessoas bilíngues (fluentes em português e em Libras), sendo uma surda e uma ouvinte (intérprete de Libras). Elas tiveram acesso ao conteúdo em Libras (glosa) traduzido pelo aplicativo e às frases originais em português e puderam contribuir com comentários sobre possíveis melhorias da solução. Foi solicitado que os indivíduos testadores realizassem a comparação entre o teor da frase original em português e o resultado da tradução e que avaliassem se a correspondência de significados, classificando assim o resultado conforme os parâmetros mencionados na seção de conceitos e tecnologias.

Além de realizar a análise das sentenças e de expor os motivos pelos quais elas não resultaram compreensíveis, os voluntários destacaram diversos pontos sobre como a Libras é uma linguagem mais objetiva que o português e ressaltaram que algumas palavras em português não possuem um sinal específico em Libras: essa fala se demonstrou alinhada

com as características de linguagem apontadas por Paiva (2019). Elas também fizeram críticas aos sistemas de legendas existentes, explicando que as *machine translation* são sensíveis a ruídos externos e a diferentes sotaques e que isso é uma grande oportunidade de melhoria no que diz respeito à qualidade na criação de legendas e inclusive na solução deste projeto.

A partir da categorização gerada nos testes, foi aplicada a métrica de eficácia SER. Ela se mostrou a métrica mais adequada para avaliar se o sentido da mensagem foi ou não devidamente mantido.

3.2.1 - Testes com algoritmo

Após a realização de testes com usuários e do cálculo do índice SER, foi realizado o teste do processo de transcrição por meio da aplicação de um algoritmo externo ao sistema. Com isso, buscava-se coletar dados a fim de compreender o motivo pelo qual algumas frases não foram adequadamente traduzidas e não puderam manter o sentido das sentenças originais. A transcrição de áudio para texto provida pela ferramenta Speech To Text da plataforma IBM Watson foi submetida isoladamente a um algoritmo em Python que calcula o índice WER. Para isso, as frases foram inseridas uma a uma no algoritmo, comparando o teor original em português com o conteúdo devolvido na transcrição, também em português, devolvendo assim o índice WER já calculado.

Com o aplicativo, foi coletada uma série de informações que forneceram ao estudo conclusões sobre a proposta de solução apresentada. Os resultados atingidos serão apresentados e discutidos na seção a seguir.

4. Resultados e discussão

Cabe lembrar que a intenção deste projeto é oferecer uma solução no que tange à tradução de conteúdos de áudios para Libras em *smartphones*. Para que essa tradução seja considerada bem-sucedida, é importante avaliar se o teor da mensagem é mantido e compreender como as etapas do sistema interferem nesse processo. Nesta seção, serão apresentados e discutidos os dados de testes realizados.

A validação foi realizada por meio da aplicação de teste com as pessoas voluntárias. O Quadro 2 apresenta as traduções e a classificação dada a elas. As expressões “pessoa surda” e “pessoa ouvinte” serão utilizadas para a classificação que demonstra respectivas análises.

Quadro 2. Resultados da etapa de testes com pessoas bilíngues

Número da frase	Frase em português	Frase traduzida para Libras pelo app	Frase incompreensível	Frase compreensível, mas com problemas	Frase compreensível
1	O bebê é muito calmo à noite.	BEBÊ CALMO_NOITE [PONTO]			Pessoa surda e pessoa ouvinte
2	Muitas motos foram roubadas na cidade.	MUITO MOTO ROUBAR CIDADE[PONTO]		Pessoa ouvinte	Pessoa surda
3	Existem muitos animais raros no zoológico.	EXISTIR MUITO ANIMAL RARO ZOOLOGICO [PONTO]		Pessoa surda e pessoa ouvinte	
4	Cuidado com o bebê e os filhotes.	CUIDADO BEBÊ FILHOTE CUIDADO BEBÊ FILHOTE CUIDADO BEBÊ [PONTO]	Pessoa surda e pessoa ouvinte		
5	Naquela rua tem muitas casas de madeira.	AQUELA RUA TER MUITA CASA MADEIRA [PONTO]		Pessoa surda e pessoa ouvinte	
6	Na loja tem muitos produtos de limpeza.	LOJA VÁRIOS PRODUTO LIMPEZA [PONTO]			Pessoa surda e pessoa ouvinte
7	O Papa Francisco é muito simpático.	PAPA&PONTÍFICE FRANCISCO ACREDITAR MUITO SIMPÁTICO [PONTO]	Pessoa surda e pessoa ouvinte		
8	Será que o vento sopra sempre na mesma direção?	VENTO SOPRAR SEMPRE IGUAL DIREÇÃO SEMPRE IGUAL DIREÇÃO	Pessoa surda e pessoa ouvinte		
9	O vento pode soprar em várias direções.	VENTO SOPRAR VÁRIAS DIREÇÃO [PONTO]	Pessoa surda e pessoa ouvinte		
10	Maria comeu muitas maçãs.	MARIA COMER(+) MANGA&FRUTA [PONTO]	Pessoa surda e pessoa ouvinte		

Fonte: Do Autor.

A pessoa surda classificou três frases como compreensíveis, duas frases como compreensíveis mas com problemas e cinco frases como incompreensíveis. A pessoa ouvinte considerou que duas frases eram compreensíveis, três frases eram compreensíveis mas apresentavam problemas e cinco frases eram incompreensíveis. Desse modo, o índice SER foi o mesmo para os dois conjuntos de dados: 0.50, ou seja, uma taxa de erro de frases de 50%.

Quando as pessoas testadoras detectaram que era possível aperfeiçoar as frases, elas descreveram características linguísticas e fizeram comentários com propostas de melhoria. Essas sugestões estão relatadas a seguir.

A pessoa surda considerou clara a tradução da frase 2 “MUITO MOTO ROUBAR CIDADE[PONTO]”, mas a pessoa ouvinte adicionaria ainda o verbo “TER” à sentença final. Em relação à frase 3, “EXISTIR MUITO ANIMAL RARO ZOOLOGICO [PONTO]”, ambos apontaram que não há um sinal para a palavra “RARO” e que a tradução mais adequada substituiria o trecho por “difícil de encontrar”.

A tradução 4, “CUIDADO BEBÊ FILHOTE CUIDADO BEBÊ FILHOTE CUIDADO BEBÊ [PONTO]”, soou confusa para ambos devido à repetição. A pessoa surda apontou que o ideal seria “cuida bebê também filhote”. A tradução tal como está parece estar se referindo a somente um bebê humano. Já para a pessoa ouvinte, o ideal seria “cuidado com o bebe e filhote animal”. A palavra “filhote” não possui uma correspondência direta em Libras, por isso, seria necessário acrescentar o sinal “animal” à tradução.

Ambas pessoas testadoras concordaram que a frase 7, “PAPA&PONTÍFICE FRANCISCO ACREDITAR MUITO SIMPÁTICO [PONTO]”, é incompreensível. A tradução parece dizer que o Papa está falando de outra pessoa, e não que outra pessoa está comentando sobre ele. Além disso, não existe um sinal para “pontífice”; o sinal para “Papa” seria suficiente.

Quanto às frases 8 e 9, respectivamente “VENTO SOPRAR SEMPRE IGUAL DIREÇÃO SEMPRE IGUAL DIREÇÃO” e “VENTO SOPRAR VÁRIAS DIREÇÃO [PONTO]”, apontou-se que as palavras “soprar” e “direção” não existem em Libras e que poderiam ser substituídas pelos sinais “vento” e “caminho”. As sugestões de traduções foram “Como vento diferente?” e “Tem vento caminho diferentes?” para a frase 8 e “vento pode diferentes caminhos” para a frase 9. Por fim, para a frase 10, “MARIA COMER(+) MANGA&FRUTA [PONTO]”, seria simplesmente substituída por “maria comer muita maçã”.

Observou-se, portanto, que 50% das frases testadas não foram compreendidas. Algumas características de linguagem mencionadas por Paiva (2019) puderam ser constatadas como falhas com relação ao consumo da API VLibras. As frases 3, 7, 8 e 9 trouxeram em sua tradução palavras que não possuem sinais em Libras; sendo as frases 7, 8 e 9 não compreendidas, ou seja, 60% das frases incompreendidas.

Em relação às frases 4 e 8, constatando-se que não houve sinais de falha na etapa de transcrição, foi possível identificar que a repetição encontrada na frase final se deu na fase de tradução da VLibras, aparentemente sem um motivo específico. Frase cujo tempo verbal era no passado (frase 2) perdeu a temporalidade após a tradução, embora não tenha sido apontado pelas pessoas voluntárias.

Na etapa de testes com algoritmo, em que ocorreu a utilização de algoritmo Python para o cálculo do índice WER, a qualidade da transcrição do áudio foi analisada de modo a se observar o impacto sobre o índice SER. O Quadro 3 apresenta o índice WER de cada frase transcrita. Ao Quadro, também foi a coluna “Tradução compreensível”, que retoma a classificação dos usuários e facilita a comparação com os demais dados.

Quadro 3. Resultados da etapa de testes computacionais e de testes com usuários

Número da frase	Frase em português	Frase transcrita com IBM Watson	Índice WER	Tradução compreensível
1	O bebê é muito calmo à noite.	bebê é muito calmo a noite	0.29	Sim
2	Muitas motos foram roubadas na cidade.	muitas motos foram roubadas na cidade	0.0	Sim
3	Existem muitos animais raros no zoológico.	existem muitos animais raros no zoológico	0.0	Sim
4	Cuidado com o bebê e os filhotes.	cuidados com o bebê e o filhote	0.43	Não
5	Naquela rua tem muitas casas de madeira.	naquela rua tem muitas casas de madeira	0.0	Sim
6	Na loja tem muitos produtos de limpeza.	na loja tem muitos produtos de limpeza	0.0	Sim
7	O Papa Francisco é muito simpático.	o papa francisco é muito simpático	0.0	Não
8	Será que o vento sopra sempre na mesma direção?	será que o vento sopra sempre na mesma direção	0.11	Não
9	O vento pode soprar em várias direções.	o vento pode soprar em várias direções	0.0	Não
10	Maria comeu muitas maçãs.	maria comeu muita manga	0.50	Não

Fonte: Do Autor.

É possível observar que, dentre as cinco frases de tradução não compreensíveis, duas possuem alta taxa de erro na transcrição (frases 4 e 10), uma apresentou taxa de erro de 11% (frase 8), e duas frases tiveram índice zero (frases 8 e 9).

Na frase 4, a transcrição trouxe a informação de que os filhotes estavam no singular. No entanto, a avaliação dos usuários apontou que a tradução perdeu significado principalmente porque não foi possível diferenciar o bebê dos filhotes. Portanto, o impacto da transcrição não se apresentou como fator principal para a perda no sentido da tradução. O contrário ocorre na frase 10: a transcrição não reconhece a palavra “maçã”, substituindo-a por “manga”, alterando assim o sentido durante o processo.

Em relação à frase 8, o WER foi de 11% e ocorreu devido a perda do sinal de interrogação no final da sentença. O que comprometeu o sentido dessa tradução foi especialmente a falta de consideração de características linguísticas no tratamento da API VLibras. As frases 7 e 9, mesmo obtendo taxa de erro de 0%, não foram compreendidas. Já a frase 1, apesar de apresentar taxa de erro de 29%, foi compreendida após a tradução; a taxa reconheceu a diferença ocorrida pela falta do artigo “o” e a ausência do acento grave em “a”.

Das 10 frases testadas, portanto, uma (10%) teve seu sentido comprometido pela transcrição realizada na IBM Watson. Considerando a totalidade dos resultados, 50% das frases não tiveram uma tradução adequada para estabelecer a comunicação desejada. 20% dessas frases foram comprometidas pela etapa de transcrição e 80% pela não observação de características de linguagem ou por repetição de palavras por parte da API VLibras. As

métricas aplicadas se mostraram adequadas ao objeto de validação: com a WER foi possível avaliar a confiabilidade da etapa de transcrição e, com a SER, foi possível verificar a eficácia da solução proposta, com a classificação e os comentários dos usuários sobre as diferenças.

5 - Considerações finais

O objetivo geral deste projeto foi tornar conteúdos de áudio acessíveis através da criação de um aplicativo. Os aplicativos de tradução para Libras mais usados não permitem a entrada de sons do *smartphone*, e o resultado deste projeto contribuiu com a possibilidade de utilização desses recursos, além de prover parâmetros para estudos de comparação, desenvolvendo uma metodologia para validação da solução proposta capaz de ser aplicada em trabalhos futuros.

Tendo a tradução gerada pelo aplicativo apresentado 50% de eficácia, concluiu-se que a API VLibras ainda não reconhece algumas particularidades da Libras, como a substituição de palavras que não possuem respectivo sinal. No entanto, não há evidências quanto ao desempenho em caso de utilização do avatar da API VLibras no lugar da glosa, o que abre espaço para a realização de trabalhos futuros responsáveis por tratar essa questão.

O índice de eficácia da transcrição foi de 90%, embora esta variável tenha sido controlada durante os testes. A avaliação do impacto da tecnologia de reconhecimento de voz no app criado também é uma proposta para próximos trabalhos. Fatores como ruído ambiente, diferentes sotaques e áudio maiores que 4 segundos podem ser estudados por meio do comparativo com os dados deste projeto.

Como o volume de dados observados nesta investigação foi limitado a duas pessoas e a 10 frases, a aplicação da metodologia de validação com mais participantes e uma maior quantidade de sentenças com estruturas diversas também poderá compor pesquisas futuras. Além disso, outros estudos poderão tratar da aplicação de tradução e transcrição via *streaming* (multimídia constantemente recebida). O código do projeto desenvolvido está disponível em <https://github.com/jaquelinebonoto/libras>.

6 - Referências

Instituto Brasileiro de Geografia e Estatística (2010). Censo demográfico 2010: Características gerais da população, religião e pessoas com deficiência. *Technical report*. Disponível em <https://censo2010.ibge.gov.br/>. Acesso em: 02 dez, 2019.

BRASIL. [Constituição (1988)]. **Constituição da República Federativa do Brasil**. Brasília, DF: Senado Federal, 2016. 496 p. Disponível em: https://www2.senado.leg.br/bdsf/bitstream/handle/id/518231/CF88_Livro_EC91_2016.pdf Acesso em: 02 dez. 2019.

COSTA, R. E. O.; ARAÚJO, T. M. U.; LIMA, M. A. C. B.; VERÍSSIMO, V. M.; ANDRADE, R. S.; VIEIRA S. F.; SANTOS, A.; SOUZA FILHO, G. L.; SOARES, M. K.; HANAEL, V. (2019). Towards an open platform for machine translation of Spoken Languages into sign languages. **SpringerNature**, Nova Zelândia Volume 33, pp 315–348. Disponível em <https://doi.org/10.1007/s10590-019-09238-5>. Acesso em: 05 nov, 2019.

PAIVA, F. A. dos S. (2019) **Tradução Automática de Português Brasileiro para Libras e Análise de Processos de Intensificação**. Tese de Doutorado. Faculdade de Engenharia Elétrica e de Computação da Universidade Estadual de Campinas.

FURLAN, A. L. (2016) **Desenvolvimento de um protótipo de aplicativo móvel para conversão de voz em texto e texto em voz, orientado ao apoio à comunicação de deficientes auditivos**. Dissertação de Mestrado. Programa de Pós-Graduação em Tecnologias da Informação e Comunicação da Universidade Federal de Santa Catarina.

TORRES, A. C. M.; ULISSES FILHO, J.; MARCOLINO, M. R. T.; DE MEDEIROS F. E.; LIMA, I. T. (2018). **Interação e comunicação da comunidade surda em grupos do aplicativo Whatsapp**. Fortaleza - Ceará. Disponível em <https://www.editorarealize.com.br/index.php/artigo/visualizar/47666> Acesso em: 13 nov, 2020.

ANDROID - Request App Permissions
<<https://developer.android.com/training/permissions/requesting>> Acesso em: 31 out, 2019.

CETIC (2018). TIC Domicílios. Disponível em
<<https://cetic.br/pesquisa/domicilios/indicadores>> Acesso em: 05 nov, 2019.

HANDTALK - Acessibilidade. Disponível em <<https://www.handtalk.me/>>. Acesso em: 05 nov, 2019.

VLibras - Download do app. Disponível em
<https://play.google.com/store/apps/details?id=com.lavid.vlibrasdroid&hl=pt_BR>. Acesso em: 13 out, 2019.

IBM WATSON - Speech to text. Disponível em
<<https://www.ibm.com/br-pt/cloud/watson-speech-to-text>>. Acesso em: 14 out, 2020.

BAIN - O impacto econômico e social do Android no Brasil. Disponível em
<<https://www.bain.com/pt-br/insights/economic-and-social-impact-of-android-in-brazil/>>. Acesso em: 17 nov, 2020.

API Vibras. Disponível em <<https://traducao2.vlibras.gov.br/docs>>. Acesso em: 31 out, 2020.