O Uso da Análise de Emoções como Auxílio na Tradução Automática de Português Brasileiro para Libras

Vinícius Matheus Veríssimo Da Silva

CENTRO DE INFORMÁTICA
UNIVERSIDADE FEDERAL DA PARAÍBA

João Pessoa, 2021

Vinícius Matheus Veríssimo Da Silva

O Uso da Análise de Emoções como Auxílio na Tradução Automática de Português Brasileiro para Libras

Dissertação de Mestrado apresentada como requisito para obtenção do título de Mestre em Informática pelo Programa de Pós-Graduação em Informática da Universidade Federal da Paraíba – UFPB

Orientador: Prof. Dr. Rostand Edson Oliveira Costa

Junho de 2021

Catalogação na publicação
Seção de Catalogação e Classificação

S586u Silva, Vinicius Matheus Verissimo da.
O uso da análise de emoções como auxílio na tradução automática de Português brasileiro para Libras / Vinicius Matheus Verissimo da Silva. - João Pessoa, 2021.
73 f. : il.

Orientação: Rostand Edson Oliveira Costa.
Dissertação (Mestrado) - UFPB/CI.

1. Informática. 2. Tradução automática. 3. Língua de sinais. 4. Análise de emoções. 5. Acessibilidade. 6. Libras. I. Costa, Rostand Edson Oliveira. II. Título.

UFPB/BC CDU 004(043)

Elaborado por Larissa Silva Oliveira de Mesquita - CRB-15/746

ATA Nº 6

Aos VINTE E NOVE dias do mês de JANEIRO do ano de DOIS MIL E VINTE E UM, às 16h, por meio de videoconferência, instalou-se a banca examinadora de dissertação de Mestrado do aluno VINICIUS MATHEUS VERISSIMO DA SILVA, matrícula 20191000933. A banca examinadora foi composta pelos professores doutores ROSTAND EDSON OLIVEIRA COSTA, UFPB (Presidente); TIAGO MARITAN UGULINO DE ARAUJO, UFPB (Interno); e VIRGINIA PINTO CAMPOS, UFRN (Externo à Instituição). Deu-se início a abertura dos trabalhos, por parte do Presidente, que, após apresentar os membros da banca examinadora e esclarecer a tramitação da defesa, solicitou ao candidato que iniciasse a apresentação da dissertação, intitulada O Uso da Análise de Emoções como Auxílio na Tradução Automática de Português Brasileiro para LIBRAS. Concluída a exposição, o Presidente, passou a palavra aos examinadores para arguirem o candidato, e, em seguida, fez suas considerações sobre o trabalho em julgamento, tendo sido APROVADO o candidato, conforme as normas vigentes na Universidade Federal da Paraíba. A versão final da dissertação deverá ser entregue ao programa, no prazo de 60 dias contendo as modificações sugeridas pela banca examinadora e constante na folha de correção anexa.

Virginia Pinto Campos

Dr. VIRGINIA PINTO CAMPOS, UFRN

Examinador Externo à Instituição

Tiago Maritan U. de Araújo

Dr. TIAGO MARITAN UGULINO DE ARAUJO, UFPB

Examinador Interno

ROSTAND EDSON OLIVEIRA COSTA, UFPB

Presidente

Vinicius Matheus V. da Silva

VINICIUS MATHEUS VERISSIMO DA SILVA

Mestrando

Universidade Federal da Paraíba

PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

FOLHA DE CORREÇÕES

ATA Nº 6

Autor: **VINICIUS MATHEUS VERISSIMO DA SILVA**

Título: **O Uso da Análise de Emoções como Auxílio na Tradução Automática de Português Brasileiro para LIBRAS**

Banca examinadora:

Prof. VIRGINIA PINTO CAMPOS	Examinador Externo à Instituição	Virginia Pinto Campos
Prof. TIAGO MARITAN UGULINO DE ARAUJO	Examinador Interno	Tiago Maritan U. de Araújo
Prof. ROSTAND EDSON OLIVEIRA COSTA	Presidente	

Os itens abaixo deverão ser modificados, conforme sugestão da banca examinadora.

COMENTÁRIOS GERAIS:

Declaro, para fins de homologação, que as modificações, sugeridas pela banca examinadora, acima mencionadas, foram aceitas e serão cumpridas integralmente.

Prof. ROSTAND EDSON OLIVEIRA COSTA

Orientador

"Do or do not. There is no try."

Master Yoda

AGRADECIMENTOS

O principal e maior agradecimento deve ser aos meus pais Valdineide Veríssimo da Silva e Valdir Félix da Silva, que puderam me proporcionar as oportunidades que me fizeram ingressar na universidade e que continuaram com o apoio para que eu pudesse concluir a graduação.

À minha amada e companheira Dyliane Mourí, por ter me apoiado, aconselhado e acalmado durante todo esse tempo juntos. Me ajudou a amadurecer não só pessoal, mas também academicamente e profissionalmente. Devo muito a ela.

Por fim, a todos os meus amigos e colegas do curso de Ciência da Computação da UFPB e do Laboratório de Aplicações de Vídeo Digital (LAVID), pela suas parcerias e trabalhos juntos, bem como aos professores que se dedicaram a proporcionar os conhecimentos que levarei para a vida.

RESUMO

Em todo o mundo a propagação de informações é voltada predominantemente para as línguas orais, isso prejudica uma parcela considerável da população, os surdos, os quais têm como língua principal a língua de sinais. Nesse sentido, as Tecnologias de Informação e Comunicação têm surgido visando minimizar essa barreira, porém ainda existem algumas críticas por parte dessa comunidade quanto ao uso dessas alternativas tecnológicas. Diante disso, esse estudo surge tendo por objetivo auxiliar as traduções automáticas de conteúdo em língua oral para língua de sinais, por meio da utilização de uma estratégia de análise de emoções em textos, empregando aprendizagem de máquina, visando apresentar um protótipo de um tradutor automático que traga maior aceitabilidade por parte dessa comunidade. Para tal, é construída uma rede neural baseada no método MultiFit para a classificação de emoções de textos em Português Brasileiro, treinada a partir de uma base de dados existente, que foi ampliada através de experimentos envolvendo *data augmentation*. Além disso, foi realizada a adaptação do tradutor automático da Suíte VLibras para que o avatar 3D da plataforma performe as emoções contidas na tradução através de expressões faciais. Apesar de inicial, a abordagem de análise de emoções apresentou uma acurácia de 94,29%, muito superior ao trabalho que originou a base de dados. Assim, os resultados são úteis por mostrarem que a adaptação do tradutor automático proporciona um ponto inicial para a adaptação de outros tradutores automáticos existentes. Além de que o estudo traz resultados promissores, os quais levam a crer que o uso do protótipo aqui desenvolvido pode vir a minimizar as críticas usuais por parte da comunidade surda ao utilizarem tradutores automáticos. Para além, os resultados aqui encontrados também colaboram para o preenchimento do *gap* existente na literatura no que se refere a pesquisas voltadas para *low-resource languages*, como é o caso da Libras.

Palavras-chave: Acessibilidade, Tradução Automática, Língua de Sinais, Análise de Emoções.

ABSTRACT

Across the world, the spread of information is predominantly directed towards oral languages, which harms a portion of the population, the deaf, whose main language is sign language. In this sense, Information and Communication Technologies have appeared to minimize this barrier, however there is still some criticism from this community regarding the use of these technological alternatives. Therefore, this study appears with the objective of helping automatic translators of content in oral language to sign language, through the use of a strategy of emotion analysis in texts, using machine learning, filling in a prototype of an automatic translator that brings more acceptability by that community. To this end, a neural network based on the MultiFit method for the classification of emotions in Brazilian Portuguese texts is built, trained from an existing database, which will be expanded through experiments involving data augmentation. In addition, the Suíte VLibras automatic translator has been adapted to that the platform's 3D avatar acts with emotions in the translation through facial expressions. Although initial, to approach to do the emotion analysis presented an accuracy of 94.29%, much higher than the work that originated the database. Thus, the results are useful because they show that the adaptation of the automatic translator offers a starting point for the adaptation of other existing automatic translators. In addition to that, the study brings promising results, which lead to believe that the use of the prototype developed here may come to minimize the usual criticism by the deaf community when using automatic translators. In addition, the results found here also collaborate to fill the gap existing in the literature regarding research focused on low-resource languages, as is the case of Libras.

Key-words: Accessibility, Machine Translation, Sign Language, Emotion Analysis

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS

LO - Língua Oral

LS – Língua de Sinais

LIBRAS – Língua Brasileira de Sinais

LSE – Língua Espanhola de Sinais

ASL – Língua Americana de Sinais

DSL - Língua Alemã de Sinais

EBMT – Tradução de Máquina Baseada em Exemplos

RBMT – Tradução de Máquina Baseada em Regras

SMT - Tradução de Máquina Estatística

LSTM - Long Short-term Memory

NLTK – Natural Language Toolkit

NN - Redes Neurais

DNN - Redes Neurais Profundas

NMT – Tradução de Máquina Neural

RNN – Rede Neural Recorrente

SVM - Support Vector Machine

KNN - K-Nearest Neighbors

CNN - Convolutional Neural Networks

QRNN - Quasi-Recurrent Neural Networks

ULMFiT - Universal Language Model Fine-tuning for Text Classification

MultiFiT - Multi-lingual Fine-Tuning

LM - Language Model

Seq2Seq - Sequence-to-Sequence

GCNLP - Google Cloud Natural Language API

IBGE - Instituto Brasileiro de Geografia e Estatística

TICs - Tecnologias e Informação e Comunicação

PLN - Processamento de Linguagem Natural

CLN - Compreensão da Linguagem Natural

GLN - Geração da Linguagem Natural

FACS - Facial Action Coding System

AU - Unidade de Ação

Conteúdo

1 INTRODUÇÃO

1.1 Contextualização

As trocas de informações que ocorrem no âmbito mundial são, majoritariamente, voltadas para os usuários de línguas orais (LO). Assim, a comunicação é feita, em grande parte, por meio de áudios, textos e vídeos voltados para pessoas ouvintes. Esse fenômeno também ocorre no Brasil, porém, uma complicação frente a esse modelo de comunicação é que uma parcela considerável da população acaba ficando à margem dele, sendo essa a população de pessoas surdas.

No Brasil, segundo o censo de 2010 do IBGE [1], por volta de 10 milhões de pessoas são consideradas surdas, o que equivale a cerca de 5% da população. Dentre essas pessoas, 2,7 milhões possuem um grau de surdez absoluta. Assim, tem-se uma relevante parcela da população que está suscetível a sofrer os efeitos de viverem em um meio onde a maioria das comunicações e distribuição de informações ocorre por meio de línguas orais.

Os surdos enfrentam uma série de barreiras que lhe são apresentadas desde os primeiros anos de vida, seja no ambiente familiar ou educacional, causando malefícios que perduram até mesmo na fase adulta [53]. Assim, diante desses obstáculos, surgiu-se uma maneira de tentar minimizar os problemas de comunicação que pessoas surdas enfrentam, sendo essa por meio do ensino de escrita de línguas orais para surdos.

A utilização do ensino da escrita de línguas orais para surdos tem sido uma tarefa de grande desafio, e nem sempre tem obtido sucesso, o que ocasiona evasão escolar de estudantes surdos, baixo rendimento em diversas áreas da grade curricular [45], e consequentemente, alta dificuldade de comunicação por essa parcela da população, visto que os surdos têm como língua natural a Língua de Sinais (LS). Assim, mesmo que ao redor da pessoa surda a LO seja a principal forma de interação, nem sempre ela estará apta a utilizar esse tipo de língua, de modo que, direta ou indiretamente, ela se utilizará de algum tipo de LS [13].

Conforme já apontado por alguns estudos [3, 54], a privação de línguas de sinais nos primeiros cinco anos de vida de uma pessoa surda dificulta a aquisição de novas línguas. Esse caso de privação de idioma é constantemente relacionado com resultados negativos nos processos de desenvolvimento cognitivo e de aprendizado [3, 54], o que impacta diretamente em como uma pessoa surda se integra e participa de sua comunidade [53], uma vez que a sua capacidade de absorver e difundir informações é limitada pelas barreiras comunicacionais.

Frente a todas as dificuldades de comunicação que as pessoas surdas enfrentam na sociedade, as áreas tecnológicas de pesquisa têm ao longo dos anos tentado realizar a

[1] https://censo2010.ibge.gov.br/

diminuição de barreiras linguísticas e de comunicação. Na verdade, essas áreas buscam esse estreitamento entre diversas línguas, e com as línguas de sinais não seria diferente, uma vez que as línguas de sinais são uma ferramenta poderosa para investigar como as línguas humanas se formam e como são processadas [17], visto que elas são estruturalmente tão complexas quanto qualquer outra língua humana [49]. Além disso, um estreitamento entre as línguas orais e as línguas de sinais é benéfico tanto para a população surda, quanto para a população ouvinte, uma vez que ambos terão um maior acesso a comunicação e troca de informações.

As áreas tecnológicas de pesquisa apresentam uma série de soluções que seriam capazes de aumentar a capacidade de comunicação entre pessoas surdas e pessoas ouvintes. Por exemplo, Bai e Bruno [3] destacam a utilização de sistemas que possibilitam a presença virtual de intérpretes de língua de sinais em uma interação onde uma das partes é uma pessoa surda, e a outra é uma pessoa ouvinte. Isso se configura como uma alternativa viável para atingir casos pontuais onde um intérprete é necessário. Todavia, essa utilização possui algumas desvantagens, como a necessidade de uma boa conexão de internet e a capacidade dos dispositivos de capturarem e reproduzirem os vídeos em uma qualidade e resolução aceitáveis, além da necessidade de disponibilidade dos intérpretes.

Assim, uma estratégia mais viável para o estreitamento de barreiras linguísticas é a utilização de algoritmos de computação visual baseados em treinamento de máquina com redes neurais profundas, como apresentados nos trabalhos de Cihan et al. [8], Ko et al. [30] e Oliveira et al. [42]. Essa técnica possibilita que vídeos de usuários de línguas de sinais transmitindo alguma informação sejam traduzidos para conteúdo em língua oral (e.g. texto). Essas tecnologias são notáveis, porque, embora ainda sejam necessários alguns recursos para a utilização das mesmas, elas independem de um intermediário humano para a realização do processo de tradução, o que aumenta a abrangência dos contextos onde elas podem ser utilizadas. Porém, esses algoritmos resolvem apenas um sentido da comunicação, traduzindo a informação em LS para informação em LO. Todavia, para realizar o processo inverso, outras estratégias semelhantes podem ser utilizadas.

Abordagens como as de Stoll et al. [62], Wairagya et al. [67] e a Suíte VLibras,de Araújo [2], que realizam a tradução de conteúdo multimídia em língua oral para língua de sinais, são uma boa alternativa para utilização não só em comunicações imediatas, mas também em casos onde nem sempre a presença de um intérprete humano é possível ou viável, como por exemplo, na internet, que possui um conteúdo extramente dinâmico e primariamente voltado para usuários ouvintes. Nesses meios, essas estratégias se destacam, pois elas se utilizam de recursos como vídeos de intérpretes, robôs e avatares 3D para transmitirem a informação em língua de sinais.

Uma visão mais concreta da utilização dessas Tecnologias de Informação e Comunicação (TICs) em um contexto mais geral pode ser encontrada no trabalho de Rocha

e Melgaço [50], que realizaram uma pequisa de opinião com usuários de aplicativos de tradução automática para Libras (Língua Brasileira de Sinais) que têm como foco a utilização de avatares 3D para performar a sinalização. No referido estudo, constatou-se que há uma visão positiva por parte da comunidade surda acerca do uso dessas tecnologias, todavia, a expressividade do avatar 3D é citado como um ponto a ser melhorado. Mais especificamente, os usuários destacaram como principais barreiras os seguintes aspectos: a interpretação do avatar, a falta de humanidade do mesmo, a necessidade de expressão de sentimentos, e a carência de expressões faciais e corporais adequadas. Entretanto, a pesquisa também apontou que os entrevistados acreditam que tais problemas podem ser sanados, de modo a elevar a aceitação dessas ferramentas dentro da comunidade surda.

Diante do exposto, surgiu-se o interesse de entender como o aprimoramento de tecnologias já existentes, por meio da consideração das emoções humanas na tradução automática para Libras, poderia impactar a capacidade de comunicação de pessoas surdas, de modo a elevar a aceitação dessas ferramentas dentro dessa comunidade. Assim, tem-se o seguinte problema de pesquisa: **Como a estratégia de análise de emoções pode auxiliar tradutores automáticos de Português Brasileiro para Língua de Sinais?**

A solução proposta utiliza uma estratégia de redes neurais baseado no método MultiFit para realizar a análise de emoções de textos em Português Brasileiro. Os resultados obtidos através dos experimentos sobre a utilização de *data augmentation* para a expansão da base de dados e aprimoramento da rede neural apontaram uma acurácia de 94,29% na classificação das emoções, para o contexto da base de dados. Além disso, também fora realizada a adaptação da ferramenta de tradução automática da Suíte VLibras para demonstrar emoções, visando trazer uma maior humanidade avatar 3D, por meio de suas expressões faciais.

O restante deste documento está organizado da seguinte forma: na Seção 1.2 são apresentados os objetivos desse estudo; no Capítulo 2 é apresentada a fundamentação teórica dos principais conceitos abordados nesse trabalho; no Capítulo 3 são apresentados trabalhos relacionados à pesquisa apresentada; no Capítulo 4 é abordada a metodologia a ser utilizada na solução proposta; no Capítulo 5 são apresentados os passos e processos realizados para a construção da solução proposta; no Capítulo 6 são apresentados e analisados os resultados desse presente estudo; por fim, o Capítulo 7 traz as conclusões acerca do trabalho apresentados, considerações finais e planejamentos para trabalhos futuros.

1.2 Objetivos

1.2.1 Objetivo Geral

Como forma de aprimorar a tradução de conteúdo em línguas orais para línguas de sinais, o trabalho tem como objetivo geral: explorar a estratégia de análise de emoções em

textos empregando aprendizagem de máquina para a utilização em tradutores automáticos de Português Brasileiro para Libras.

1.2.2 Objetivos Específicos

Para que a solução proposta tenha êxito, faz-se necessário, portanto, que os seguintes objetivos específicos sejam atingidos:

- Explorar a construção de um sistema baseado em aprendizagem de máquina para atribuição de emoções em texto utilizando bases de dados existentes;
- Realizar experimentos para a ampliação da base de dados de treinamento do algoritmo de aprendizagem de máquina utilizando *data augmentation*;
- Utilizar o sistema desenvolvido em conjunto com a adaptação de uma ferramenta de tradução automática de língua oral para língua de sinais.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 Processamento de Linguagem Natural

O Processamento de Linguagem Natural (PLN) é uma sub-área da Ciência da Computação, responsável por utilizar técnicas computacionais para aprender, entender e produzir conteúdo em linguagem humana [23]. Esse campo de estudo está relacionado com diferentes teorias e técnicas que lidam com problemas relacionados à utilização de linguagem natural na comunicação com computadores [29]. O PLN pode ser utilizado para diversos propósitos, tais como: tradução automática, reconhecimento de fala, síntese de texto, análise de sentimento, sumarização, entre outros.

Segundo Khurana et al. [29], o PLN pode ser classificado em duas partes: Compreensão da Linguagem Natural (do inglês, *Natural Language Understanding*) (CLN) e Geração da Linguagem Natural (do inglês, *Natural Language Generation*) (GLN). A parte de CLN está mais envolvida com a utilização de conceitos linguísticos como fonologia; que se refere ao som; morfologia, referente à formação das palavras; sintaxe, referente à estrutura das sentenças; semântica, relacionada com o significado do texto; e pragmatismo, relacionado ao uso real da linguagem. Além disso, ela ajuda a explicar significados intrínsecos do texto que não são explicitamente descritos, mas estão compostos no contexto. Por outro lado, a parte de GLN é o processo de produzir frases, sentenças e textos a partir de uma representação interna significativa, que em suma realiza um procedimento oposto à CLN.

A literatura acerca do PLN tem se desenvolvido nos últimos anos, todavia, um empecilho para um maior avanço da área são os diversos desafios que os processos de compreensão e geração de linguagens naturais apresentam. Dentre esses, um desafio bem conhecido é a ambiguidade, que é a capacidade de uma palavra ou texto ter mais de uma interpretação válida. Assim, um meio para resolver esse problema em específico seria a obtenção de mais conhecimento sobre a linguagem, o texto ou o contexto, dependendo do tipo de ambiguidade.

Por outro lado, Goldberg [20] identifica três propriedades nas linguagens naturais que são desafiadoras paras os sistemas computacionais, sendo essas:

- Quando são simbólicas e discretas - As linguagens utilizam símbolos que denotam objetos, conceitos, eventos, ações e ideias. Caracteres e palavras são símbolos discretos e distintos que possuem significados externos a eles, os quais estão na cabeça de quem os interpreta.
- Quando são composicionais - Letras são utilizadas para formarem palavras, e palavras para formarem frases e sentenças. Quanto maior a composição, mais complexo

é o significado e a interpretação do texto. Além disso, as regras da língua ditam como as composições devem ser realizadas [29].

- Quando possuem escassez de dados - A forma como os símbolos discretos podem ser combinados para gerarem significados é praticamente infinita. Em outras palavras, é impossível mensurar a quantidade de sentenças válidas que podem ser geradas pela composição dos símbolos de uma língua. Assim, nenhum sistema é capaz de ser genérico o suficiente para uma língua.

Hirschberg e Manning [23] apontam algumas outras dificuldades enfrentadas pelo PLN. Para os autores, os recursos e sistemas disponíveis são majoritariamente direcionados para os idiomas de altos recursos (do inglês, *high-resource languages*), como Inglês, Francês, Espanhol, Alemão e Chinês. Em contraponto, os idiomas com poucos recursos (do inglês, *low-resource languages*), que são falados e escritos por milhões de pessoas, não possuem tantos dados e sistemas disponíveis, o que dificulta a construção de novas soluções voltadas para essas línguas.

Dessa forma, nota-se que esses idiomas de altos recursos possuem mais pesquisas direcionadas, ferramentas de suporte, bases de dados e produtos que os demais. Havendo assim uma dificuldade quanto à pesquisa e desenvolvimentos de soluções voltadas às línguas com poucos recursos, como é o caso das línguas de sinais.

Nesse sentido, a próxima seção aborda um pouco sobre as complexidades do processamento das línguas de sinais.

2.2 Processamento de Línguas de Sinais

Embora as línguas orais e as línguas de sinais possam transmitir a mesma informação, elas se diferenciam na forma de realizar essa transmissão, visto que uma é auditivo-oral, e a outra é visual-espacial (ou gesto-visuais), respectivamente [11]. Nas línguas orais, a garganta, nariz e boca são utilizadas como articuladores, enquanto que nas línguas de sinais se utilizam dedos, mãos, braços e expressões faciais, além de movimento, posição e espaço de sinalização [1].

As LS são tão complexas quanto qualquer outra língua natural humana, pois elas possuem variações linguísticas não só entre comunidades de usuários, mas também quanto aos aspectos fonológico, morfológico, sintático e semântico-pragmático [49]. Em outras palavras, ASL (*American Sign Language*), Libras (Língua Brasileira de Sinais), LSE (*Lengua de Signos Española*), e muitos outros exemplos de LS não compartilham as mesmas regras gramaticais, e nem o mesmo dicionário de sinais, pois se baseiam nas línguas orais utilizadas pela sociedade que as criaram, e por todo o contexto que as envolvem.

O processamento das línguas de sinais pode se dar pela interpretação de imagens e vídeos da execução dos gestos, ou pela geração de informação em alguma língua de sinais a partir de um conteúdo em alguma língua oral, ou seja, traduzir informação em LO para LS. Em ambos os contextos é comum que em algum momento se faça a representação da informação em uma descrição textual da língua de sinais, às vezes como procedimento parcial ou final da tradução.

As principais formas de representar textualmente uma língua de sinais são:

- **Notação de Stokoe** - Stokoe foi o responsável por iniciar os estudos sobre as LS. Em seus trabalhos, ele propôs que os sinais eram compostos de três partes (ou parâmetros), sendo eles: localização do sinal, forma da mão e movimento. Palma da mão e sinais não-manuais são tratados indiretamente no sistema de Stokoe [66]. Stokoe chamou esses parâmetros de *quiremas*, semelhantes aos fonemas nas LO, que são os menores elementos sonoros que permitem distinguir o significado em palavras. Cada parâmetro possui um conjunto de símbolos para designar suas variações, sendo compostos por letras, números e símbolos especiais.

  Na Figura 1 é apresentada a transcrição da palavra *IDEA* (inglês de "ideia") em ASL, na notação de Stokoe, onde: $\cap$ indica a localização na testa, | representa o formato da mão e $\wedge$ indica o movimento ascendente [66].

Figura 1: Sinal *IDEA* na notação de Stokoe [66]

- **HamNoSys** - O nome HamNoSys é um acrônimo para *Hamburg Notation System*. Essa notação foi desenvolvida pelo Grupo de Pesquisa da Universidade de Hamburgo e tem como base os parâmetros da notação proposta por Stokoe. O HamNoSys é uma notação simbólica e capaz de expressar o formato da mão, a orientação da mão, a localização, as ações, a utilização de duas mãos e os componentes não-manuais [21].

  A Figura 3, por exemplo, apresenta a escrita de *HOUSE* (inglês de "casa") em ASL na notação HamNoSys.

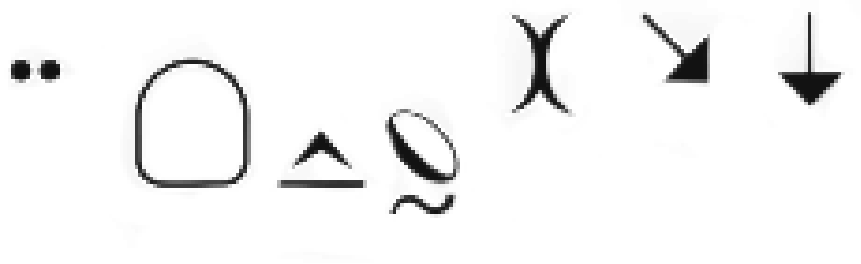

Figura 2: Descrição da sinalização de *HOUSE* na notação HamNoSys [22]

- **SignWriting**[2] - SignWriting é um sistema prático de escrita para línguas de sinais, composto por um conjunto intuitivo de símbolos gráfico-semânticos e de regras simples para serem combinadas com os símbolos visando a representação dos sinais. Diferentemente do HamNoSys, ele não é um sistema para ser utilizado por linguistas em suas representações analíticas de sinais. O SignWriting é essencialmente um sistema concebido para ser utilizado por surdos no seu dia a dia, para os mesmos propósitos que ouvintes utilizam escritas de línguas orais. Assim como as demais representações simbólicas, o sistema se baseia em representar os principais aspectos utilizados nas LS, como a configuração das mãos, o movimento das mãos e dos dedos, a localização, a expressão facial etc [51].

  A Figura 3 traz um exemplo de escrita da palavra *BRASIL* na notação SignWriting.

Figura 3: Escrita de *BRASIL* na notação SignWriting

- **Glosa** - A glosa é simplesmente a grafia do significado do sinal na língua oral na qual a língua de sinais se baseia, geralmente escrita com letras maiúsculas. Ainda podem ser acrescentadas algumas marcações não-manuais e usos do espaço de sinalização, que são representados por letras ou números subscritos [35]. A glosa respeita as regras sintáticas e semânticas da LS a qual está associada. Apesar de ser facilmente interpretada até por quem não é usuário da LS, a glosa não possui um padrão de escrita definido. É comum a modificação dela para a representação dos parâmetros

[2] http://www.signwriting.org/

das LS em texto corrido. A glosa é a representação que mais se aproxima da escrita das LO.

Na Figura 4 é apresentado um exemplo de escrita da frase *"Eu nunca vou à casa dele"* na notação glosa-Libras. Ou seja, a frase é escrita em Português Brasileiro, idioma no qual a Libras se baseia, mas respeitando as regras gramaticais da Libras.

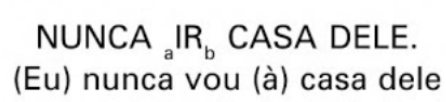

Figura 4: Exemplo da representação em glosa da frase "Eu nunca vou à casa dele"[48]

A partir dessas representações textuais que foram supracitadas é possível, computacionalmente, gerar a informação em alguma língua de sinais, seja por meio de símbolos especiais, pela execução de vídeos referentes aos sinais, ou pela sinalização através de um avatar 3D animado. Alguns exemplos desses sistemas serão apresentados no próximo capítulo.

Na próxima seção serão apresentados os principais conceitos da análise de emoções e como ela pode ser realizada de forma automática em textos por meio da utilização de redes neurais.

2.3 Análise de Emoções

Conforme apresentado por Rodrigues e Rocha [52] e Coppin e Sander [10], as emoções podem ser definidas de diversas maneiras, de tal modo que não há um consenso na literatura sobre o que são as emoções. Dessa forma, cada área de estudo tem seus próprios conjuntos de conceitos.

Coppin e Sander [10] são favoráveis à definição de Sander [57], de que a emoção é um processo rápido de dois passos. Assim, a emoção se baseia em eventos, que consiste em (1) um mecanismo de elicitação emocional baseado em relevância que (2) molda uma resposta multi-emocional. Nesse sentido, a resposta pode ser, por exemplo, uma tendência de ação, uma reação automática, expressão ou sentimento.

De forma semelhante, Niedenthal e Ric [40] definem emoções como o fogo que alimenta o comportamento humano e as forças motivadoras da vida. Logo, exemplificando, o processo de sentir uma emoção como o *medo* requer a percepção de algo no mundo, lembrar que aquilo é uma ameaça e realizar uma ação de escape.

Como tentativa de modelar as emoções humanas, Plutchik [47] criou um modelo que identifica 8 emoções primárias, sendo essas: raiva, antecipação, alegria, confiança,

medo, surpresa, tristeza e desgosto. Na Figura 5 é possível ver o modelo representado graficamente por um conjunto de círculos circunvizinhos, de modo que o círculo do meio é composto pelas emoções primárias, o círculo interno é composto pela maior intensidade das emoções primárias e o círculo externo é composto pela menor intensidade das emoções primárias. As demais emoções são misturas das emoções primárias.

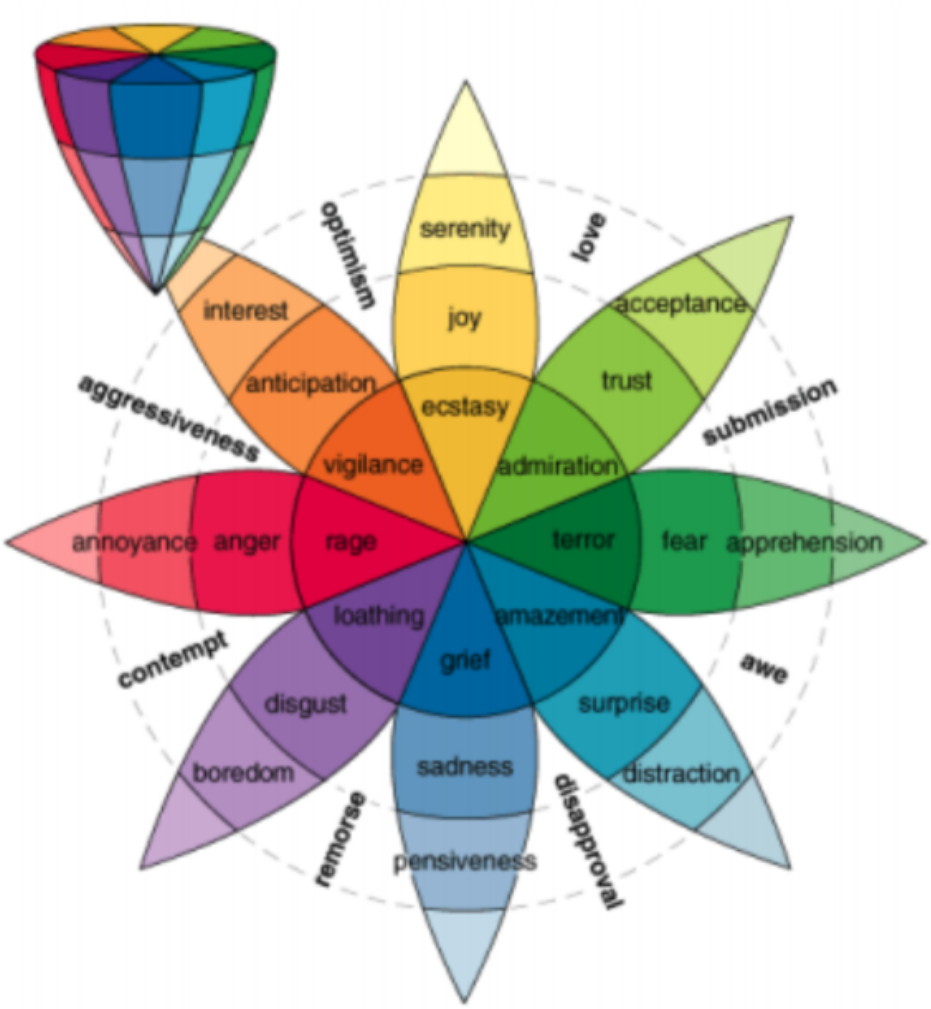

Figura 5: Círculo de emoções proposto por Plutchik [47]

De modo a entender melhor o modelo de Plutchik, representado na Figura 5, cita-se um exemplo, onde a emoção de alegria (*joy*), em maior intensidade, se manifesta como êxtase (*ecstasy*), e em menor intensidade, como serenidade (*serenity*), e quando misturada com confiança (*trust*) se manifesta como amor (*love*). A representação se assemelha à da roda de cores, de tal forma que as emoções opostas estão em um ângulo de 180º como cores complementares. Além disso, as intensidades das emoções são identificadas nas intensidades das cores. Na Figura 5 também está presente uma representação tridimensional da

modelo que ajuda na visualização das intensidades das emoções.

Para além de Plutchik [47], existe uma série de autores que também buscaram detectar e analisar as emoções humanas. Nesse sentido, a área de detecção/análise de emoções, por vezes também referenciada como detecção/análise de sentimentos ou mineração de opinião (do inglês, *opinion mining*), é o estudo de detectar e investigar emoções. Essa área tem sido explorada há anos pelas ciências humanas e biológicas, e mais recentemente pela Ciência da Computação, muito devido ao aprimoramento do poder computacional e ao aumento do uso da comunicação por meio de sistemas eletrônicos. Por isso, têm-se dado muita atenção à análise de emoções de forma automática [65]. Essa análise de emoções é essencial na Computação Afetiva, que tem o objetivo de desenvolver sistemas computacionais que possam entender e reagir aos estados afetivos do usuário [43].

Embora a análise de sentimentos e a análise de emoções sejam citadas diversas vezes como sinônimos, alguns autores as diferenciam [26, 43, 55]. Para tais autores, a análise de sentimentos é apenas o estudo da polaridade do sentimento (positivo ou negativo), já a análise de emoções abrange diferentes classes de emoções (e.g. alegria, tristeza, medo, etc...), podendo até envolver a identificação da intensidade dessas emoções. Nesse sentido, pode-se dizer que a análise de emoções é significativamente uma tarefa mais complexa que a análise do sentimento [55].

A detecção/análise de emoções pode ser realizada por meio de diferentes tipos de dados, como imagem, vídeo, áudio e texto, que podem ser utilizados em procedimentos automatizados ou semi-automatizados [55]. Esse tipo de análise em textos pode parecer simples a priori, mas a dimensionalidade das línguas humanas torna essa tarefa extremamente complexa. Além de informações, textos também são capazes de expressar a opinião e emoção do autor [14], todavia, na maior parte do tempo as emoções são expressas de forma subentendida. Assim, até mesmo uma pequena porção de texto pode conter uma série de emoções, sendo capazes de serem interpretadas de diversas formas distintas, dada a ambiguidade da língua e o conhecimento de quem a está interpretando [55].

A detecção/análise de sentimentos em texto podem ocorrer em três níveis, a saber:

- **Nível de Documento** - Nesse nível a classificação do sentimento é atribuída a todo o documento [60]. Por exemplo, um documento contendo avaliações de um produto expressaria o sentimento geral acerca daquele produto. Essa estratégia não tem aplicabilidade quando um documento possui múltiplas entidades, como por exemplo, avaliações de produtos distintos [26].

- **Nível de Sentença** - Nesse nível a tarefa de classificação de sentimento ou emoção é realiza para cada sentença [26]. Dessa maneira, cada sentença é considera um unidade separada que deve conter apenas um sentimento/emoção/opinião [27].

- **Nível de Aspectos** - Esse nível também é conhecido como Nível de Característica (do inglês, *Feature Level*). Ele é direcionado para contextos em que se é necessária uma análise em uma granularidade menor. Nesse nível se busca a análise de um termo específico, e não do documento, sentença ou frase [26]. De forma geral, nesse nível é necessário definir o aspecto que será analisado, rastrear todo os conteúdos relacionados a ele no texto e classificar o sentimento/emoção/opinião desses conteúdos [60]. Esse é o mais complexo dos três níveis.

A automação da análise de emoções em textos pode ser explorada de diversas formas, não só como parte do aprimoramento da interação homem-máquina. Áreas como mídias-sociais, negócios, medicina, entre outras, têm muito a se beneficiarem desse ramo de pesquisa. Por isso, existem diversas estratégias para diferentes contextos de uso da análise de sentimentos e emoções. Apesar da complexidade da tarefa, técnicas de aprendizagem de máquina podem ser utilizadas na construção dessas soluções [14]. Algumas dessas técnicas são:

- **Naïve Bayes** - É um algoritmo de aprendizagem probabilística [38]. Esse algoritmo se baseia no Teorema de Bayes, onde todos os termos são independentes, cenário que nem sempre reflete o mundo real [58]. Nessa solução, o processo se dá pelo cálculo da probabilidade de cada termo de um conjunto, como por exemplo palavras num documento ou sentença, pertencer a uma dada classe. A partir disso, calcula-se a qual classe é mais provável do conjunto pertencer. Essas probabilidades são calculadas a partir dos dados de treinamento [38].

- ***Support Vector Machine*** - A Máquina de Vetores de Suporte (SVM) (do inglês, *Support Vector Machine*) é um tipo de algoritmo de aprendizagem de máquina para a realização de treinamento supervisionado. Essa técnica tem uma forte base na Teoria do Aprendizado Estatístico e foi utilizada em diversas aplicações devido a sua precisão. Em um primeiro momento essa solução havia sido projetada para o aprendizado linear, possibilitando a utilização em classificações binárias. Todavia, em casos onde as classes não podem ser linearmente separadas, é possível modificar o SVM ou os dados de entrada. [14, 38]

- ***K-Nearest Neighbors*** - É um algoritmo que trabalha levando os dados para o plano vetorial e calculando a categoria de um elemento considerando a categoria dos seus k vizinhos mais próximos [4]. Os vizinhos podem ser descobertos utilizando qualquer tipo de método para cálculo de distância vetorial.

- **Redes Neurais** - Apesar de ser uma técnica tão antiga ou mais que as anteriores, ela atraiu o interesse de diversas áreas de pesquisa, principalmente na última década.

As redes neurais (NNs - do inglês, *Neural Networks*) são um tipo de modelo computacional para a realização de tarefas de reconhecimento de padrões. Além disso, elas são inspiradas na estrutura e no funcionamento das redes neurais biológicas humanas [69]. Normalmente, elas são representadas como um conjunto de nós conectados, arranjados em forma de rede e agrupados em camadas, geralmente divididas em camadas de entrada, ocultas e de saída. Cada conexão de nós está associada a um peso, que é calculado durante a fase de treinamento [38]. Dependendo da dimensionalidade dessas camadas, essas redes recebem a nomenclatura de Redes Neurais Profundas (DNN - do inglês, *Deep Neural Networks*)

Frente ao reconhecimento da importância e do uso das Redes Neurais nas últimas décadas, a próxima seção aborda como a estratégia de Redes Neurais é utilizada em tarefas de análise de emoções em textos.

2.3.1 Análise de Emoções em Texto com Redes Neurais

O uso de Redes Neurais representa uma poderosa e atrativa ferramenta para o PLN [20], visto que muitos modelos de redes neurais atingem ou até mesmo superam o estado da arte em tarefas do PLN quando comparados com outras estratégias de aprendizagem de máquina [32].

Um componente das *Neural Networks* que possibilitou a utilização de textos foi a camada de *embedding* [20]. Essa camada é responsável por mapear os símbolos contínuos (letras ou palavras) em valores matemáticos contínuos de baixa dimensionalidade. Sobre esses valores serão aplicadas as operações pelas camadas da rede [32, 20]. De forma geral, cada letra ou palavra do texto será representada como um vetor de valores reais.

A tarefa de mapear palavras em vetores pode ser realizada de uma forma mais simples utilizando a técnica de *one-hot*, onde cada palavra é representada por um vetor binário caracterizado por conter o valor 1 em apenas uma posição, com as demais posições recebendo o valor 0, de modo que cada palavra do contexto trabalhado é mapeada para um vetor distinto. Uma outra forma mais complexa de realizar esse mapeamento é gerar vetores de forma semântica, ou seja, o vetor de uma palavra é definido também levando em consideração seu uso na língua. Para tal abordagem é necessário criar ou utilizar um sistema específico, como o GloVe [46] ou o Skip-gram [36].

Exitem dois tipos de redes neurais que são majoritariamente utilizadas no processamento de linguagens naturais, são elas: Redes *Feed-forward* e Redes Recorrentes (RNN - do inglês, *Recurrent Neural Networks*) [20].

As Redes *Feed-forward* são o tipo mais comum de NN. Elas se caracterizam por organizarem os neurônios em camadas, essencialmente os organizando em camada de

entrada, camadas ocultas e camada de saída [18]. Um exemplo bem comum desse tipo de NN são as Redes Convolucionais (CNN - do inglês, *Convolutional Neural Networks*), que são compostas de várias camadas de convoluções com funções de ativação não lineares aplicadas sobre a camada de entrada, o que resulta em conexões locais capazes de localizar padrões locais. Além disso, cada camada pode aplicar diferentes tipos de filtros nos dados [34]. As CNNs são amplamente utilizadas em diversas áreas, como: reconhecimento de fala, visão computacional, PLN, etc[43].

As Redes Recorrentes foram feitas para trabalharem com informações sequenciais, elas são chamadas de recorrentes por aplicarem o mesmo procedimento em todos os elementos da sequência de entrada. As RNNs são conhecidas pelo seu mecanismo de memória, em que os elementos anteriores já computados da sequência também influenciam na computação dos próximos elementos [34]. De forma geral, a sequência recebida pela RNN pode conter um tamanho variável e a rede produzirá um vetor de tamanho fixo que sumariza toda a informação da sequência [20].

Um texto pode ser intuitivamente visualizado como um sequência de *tokens* (letras ou palavras) com dependência temporal, ou seja, um *token* na posição t pode influenciar o sentido de um outro *token* em $t + n$, assim como pode ser influenciado por um *token* em $t - n$. As CNNs são capazes de extrair informações dos textos em múltiplos níveis, possibilitando que essas redes manipulem as dependências locais e distantes [43, 9]. O tipo mais básico das RNNs não consegue ter a mesma performance das CNNs. Todavia, um caso especial são as redes LSTM (*Long-Short Term Memory* [24], um tipo de RNN que possui a capacidade de aprender dados com dependências temporais de curto e longo alcance [64].

Em suma, as RNNs integram informações de contexto atualizando um estado oculto a cada etapa de tempo, enquanto que as CNNs resumem um contexto de tamanho fixo em várias camadas [68].

Tão importante quanto os tipos de redes neurais utilizadas na construção dos sistemas é a forma como essas redes estão arquiteturalmente definidas quanto ao tamanho, quantidade e densidade das camadas, e qual método de treinamento utilizado. Uma parte considerável dos trabalhos sobre *Deep Neural Netwoks* (DNN) exploram exatamente isso, em busca de alternativas de redes neurais e métodos de treinamento que sirvam a um propósito específico ou geral.

Muitos dos problemas da área de PLN fazem parte de um conjunto maior de problemas que são o foco dessas soluções. Por exemplo, a arquitetura *Encoder-Decoder* [7] é bastante conhecida e aplicada em problemas *Sequence-to-Sequence*, como tradução, sumarização, reposta de perguntas, etc. Já a análise de sentimentos/emoções pode ser tratada como um problema de classificação de textos, onde as classes dos textos são as

próprias emoções. Nesse caso, métodos destinados para essa classe de problema também podem ser utilizadas para esse propósito.

Existem muitas formas de se utilizar DNN na análise de sentimentos e emoções em texto [12, 63, 28, 6]. Nesse sentido, muitos dos trabalhos desenvolvidos nessa área foram ou chegaram perto de atingir o estado da arte para essa tarefa, e mostram a eficiência na utilização de redes neurais para a análise de sentimentos e emoções, mesmo que nem sempre as soluções sejam diretamente projetadas para isso, mas a redução desse problema para uma tarefa de classificação de textos possibilita esses resultados.

Redes Neurais mais diretamente relacionadas com o presente trabalho serão apresentadas no próximo capítulo.

3 TRABALHOS RELACIONADOS

Nesse capítulo são apresentados trabalhos relacionados aos escopos de análise de emoções em texto, tradução automática de conteúdo em língua oral para língua de sinais e trabalhos relacionados à representação de emoções em línguas de sinais de forma natural e computacional.

3.1 Análise de Emoções

Os trabalhos subsequentes apresentam arquiteturas de classificação usando redes neurais e outras estratégias de aprendizagem de máquina para a realização da classificação de emoções e sentimentos em diferentes contextos.

No trabalho de Souza [61], a análise de sentimentos é aplicada sobre *tweets* a fim de relacionar o sentimento expresso por investidores sobre as empresas, com os retornos e volumes diários das mesmas no mercado acionário brasileiro. Para a atribuição dos sentimentos foi utilizado o *Google Cloud Natural Language API* (GCNLP), uma API em nuvem do Google para PLN. A GCNLP atribui o sentimento aos textos por meio de valores reais entre -1 (texto com sentimento totalmente negativo) e 1 (texto com sentimento totalmente positivo). A GCNLP utiliza uma estratégia com base em aprendizagem de máquina, todavia, ela não possui um acesso totalmente gratuito.

Howard e Ruder [25] propõem um método de transferência de aprendizado que pode ser aplicada para qualquer tipo de tarefa de processamento de linguagem natural. O método ULMFiT (*Universal Language Model Finetuning*) consiste em utilizar uma rede neural para *Language Model* (LM), que é capaz de compreender as dependências das palavras da língua que está sendo trabalhada, diminuindo a complexidade necessária para a construção da rede responsável pelo processo de PLN mais específico.

O método ULMFiT (*Universal Language Model Finetuning*), utilizado no trabalho de Howard e Ruder [25], consiste em utilizar uma rede LM pré-treinada para um contexto geral, especializá-la para o contexto que está sendo trabalhado, e utilizar o conhecimento adquirido em uma rede especializada para a tarefa de PLN desejada através de *transfer learning*. Como experimento, o método proposto foi utilizado em tarefas de classificação de textos (análise de sentimentos, classificação de questões e classificação de tópicos). A proposta conseguiu atingir o estado-da-arte para as bases de dados utilizadas. Esse método serviu como base para o método MultiFiT.

Eisenschlos [15] se baseia no ULMFiT para a definição de um novo método de *transfer learning* para tarefas de PLN. Apesar do ULMFiT ter sido pensado para ser utilizado com qualquer língua natural, o MultiFit aperfeiçoa isso levando em consideração que nem todas as línguas possuem construção gramaticais semelhantes ao Inglês. Sendo

assim, enquanto que o ULMFiT limita os dados de entrada à granularidade de palavras, o MultiFit possibilita a utilização de sub-palavras. Outra diferença está na utilização de redes QRNN (Quasi-Recurrent Neural Networks) em todo o processo, em vez das AWD-LSTM, as quais são utilizadas no ULMFiT, o que possibilita mais paralelização do treinamentos das redes. Além disso, as QRNN são mais leves e rápidas que as redes AWD-LSTM. Quanto aos resultados, é possível verificar que o MultiFiT supera o ULMFiT nas tarefas de classificação, principalmente quando comparados na utilização com diferentes tipos de línguas. Diante disso, esse método servirá como base para o processo de classificação utilizado na presente pesquisa.

Omara, Mosa e Ismail [43] estudaram a utilização de redes neurais convolucionais treinadas utilizando *transfer learning* a fim de realizar detecção de emoções de textos em árabe. Sobre a base de dados utilizada foi aplicada uma técnica de *data augmentation* através da substituição de sinônimos com o propósito de deixar a rede neural mais robusta. A estratégia proposta atingiu o estado da arte para a classificação de emoções em textos em árabe. Esse trabalho se relaciona à pesquisa em demonstrar a utilização de *data augmentation* no treinamento de uma rede neural que se utiliza de *transfer learning*, que compõe um dos passos dessa pesquisa.

O artigo de Dosciatti, Ferreira e Paraiso [14] utiliza Máquina de Vetores de Suporte (SVM) na identificação de seis emoções básicas (alegria, tristeza, raiva, medo, desgosto e surpresa), além da emoção neutra, em textos escritos em Português Brasileiro. Para testar o método proposto, um corpus formado por 1.750 textos foi construído. Destaca-se ainda que o método não utiliza nenhum recurso linguístico adicional, como um léxico especialmente preparado para relacionar palavras ligadas às emoções. A base de dados utilizada no presente estudo será a mesma utilizada nesse trabalho [14].

3.2 Tradução Automática de Língua Oral para Língua de Sinais

Os trabalhos envolvendo tradutores automáticos de LO para LS que serão apresentados a seguir demostram como alguns sistemas de tradução possuem partes semelhantes (e.g. tradutor automático e sinalizador automático), que são bem divididas, mesmo se utilizando diferentes estratégias de implementação.

San-Segundo et al. [56] apresentam em seu trabalho um sistema capaz de traduzir Espanhol em LSE. O sistema é dividido em três módulos: o primeiro é responsável por transcrever o áudio em texto em espanhol; o segundo módulo é responsável por traduzir o texto em espanhol para glosa-LSE; e o último módulo realiza a sinalização da glosa por meio de um avatar 3D. O tradutor automático de texto para glosa empregado no trabalho se destaca por utilizar um forma de Tradução Automática híbrida, com as três técnicas clássicas de tradução automática: a Tradução Baseada em Exemplos (EBMT),

a Tradução Baseada em Regras (RBMT), e a Tradução Estatística (SMT).

Conforme San-Segundo et al. [56], a Tradução Baseada em Exemplos (EBMT - do inglês, *Example-based Machine Translation*) realiza a tradução a partir de uma base de dados de traduções, escolhendo a mais provável de ser a correta; a Tradução Baseada em Regras (RBMT - do inglês, *Rule-based Machine Translation*) utiliza um conjunto de regras que descrevem como realizar a tradução entre as línguas; a Tradução Estatística (SMT - do inglês, *Statistical Machine Translation*) utiliza modelos estatísticos para realizar a tradução. Assim, em seu estudo, os autores obtiveram destaque ao fazerem uso de um sistema que utilizava uma forma de Tradução Automática Híbrida, com essas três técnicas citadas (EBMT, RBMT e SMT). Todavia, embora esse sistema tenha apresentado baixo erro nas métricas computacionais, ele não apresentou bons resultados quando testado num contexto com usuários reais.

Em seu estudo, Gago et al. [19] exploram a tradução de texto em língua oral para língua de sinais sendo performadas por um robô humanoide. O sistema utiliza uma DNN do tipo Sequence-to-Sequence (Seq2Seq) com redes LSTM para realizar a tradução de textos em espanhol para glosa-LSE. A sinalização da tradução em glosa-LSE é performada por um robô humanoide que acessa uma base de dados contendo os movimentos de cada palavra. Os movimentos foram obtidos pela captura do movimento de pontos de interesse de intérpretes de LSE gesticulando as palavras. Assim, a partir disso foi realizada uma representação em 3D dos movimentos. Esse estudo não apresentou resultados relacionados a testes com usuários, de modo que o mesmo carece de uma avaliação de caso de uso, visando verificar se o que fora desenvolvido é útil em um contexto com usuários reais.

Ainda no que tange ao desenvolvimento de soluções de tradução, a Suíte VLibras [2] é um conjunto de ferramentas para a tradução de conteúdo multimídia em Português Brasileiro para Libras, sinalizada por um avatar 3D. O sistema é composto de programas para computadores pessoais, aplicativos móveis, *plug-in web*, serviços de nuvem, entre outros. Esse sistema pode ser utilizado em diversas esferas, sendo ele atualmente utilizado até mesmo por alguns sites públicos governamentais. A proposta inicial de Araújo [2] para o sistema contava com um tradutor automático baseado em regras, porém, trabalhos recentes [41] têm explorado a modificação total para um tradutor automático neural com uma DNN do tipo Seq2Seq com redes convolucionais. A estratégia de sinalização automática por avatar 3D utilizada na Suíte VLibras possui uma boa recepção por parte de usuários surdos desde sua versão inicial, fato que é comprovado por sua expansão diversa.

3.3 Representação de Emoções em Línguas de Sinais

Os trabalhos a seguir abordam estudos que têm como foco emoções em línguas de sinais, tanto a partir da forma natural, por meio da análise das expressões e movimentos de um intérprete humano, quanto a partir de uma abordagem computacional, por meio da representação das emoções das línguas de sinais em sinalizadores automáticos.

No trabalho de Lemos [31], em determinado ponto, são apresentadas estratégias para a representação de efeitos sonoros e emoções em traduções automáticas de conteúdo em Português Brasileiro para Libras, no contexto de TV Digital. A representação dessas informações foi realizada na forma de pistas visuais. Para o caso do efeitos sonoros, são apresentadas imagens iconográficas relacionados ao efeito sonoro, como por exemplo, a imagem de um telefone para som de telefone ou notas musicais para indicar uma música de fundo. Para a representação de emoções são utilizados *emoticons* e cores diferencias para indicar os distintos interlocutores em cena, bem como sua emoção.

Silva et al. [59] exploram a criação de uma rede neural capaz de codificar as expressões faciais presentes na Libras em termos de unidades de ação (AU - do inglês, *action units*) do sistema FACS (*Facial Action Coding System*) de Ekman e Friesen [16]. O foco do trabalho é explorar a identificação das Expressões Faciais de Sentenças (questão, afirmação e negação) e de Expressões Faciais Afetivas (emoções), utilizando redes neurais baseadas em redes CNN e CNN+LSTM. Para o treinamento, utilizou-se uma base de dados de vídeos de intérpretes de Libras retirados da internet, anotados pelos próprios autores. Apesar das limitações da base de dados e do algoritmo, os autores consideraram a performance do modelo aceitável. A tabela de descrição das expressões faciais presentes nesse trabalho serviu como base para a definição das expressões faciais adotadas para a representação de cada emoção de forma automática na presente pesquisa.

Frente ao levantamento da literatura e ao que fora utilizado nos trabalhos relacionados, o presente estudo segue o próximo capítulo com a definição do método utilizado para atingir os seus objetivos gerais e específicos.

4 PROPOSTA

Esse estudo tem como objetivo explorar a estratégia de análise de emoções em textos empregando aprendizagem de máquina para a utilização em tradutores automáticos de Português Brasileiro para Libras.

Para a utilização da análise de emoções em conjunto com uma ferramenta de tradução automática de língua oral para língua de sinais, faz-se necessária a adaptação de alguns módulos dessa ferramenta.

A adaptação a ser apresentada terá como base uma ferramenta de tradução automática de Português Brasileiro para Libras, que performa a sinalização por meio de um avatar 3D. Nessa seção serão tratados os processos de adaptação da ferramenta de tradução automática de forma genérica.

Para a adaptação da ferramenta de tradução automática são realizados três procedimentos: adaptação do módulo de tradução automática para a aplicação da análise de emoções; adição da emoção identificada à tradução; e adaptação do sinalizador automático com avatar 3D.

Nas próximas seções serão apresentados os conceitos e definições para esses três procedimentos, tendo como foco a adaptação de um tradutor automático de texto para glosa, a sintaxe para adição de emoções a essa glosa e a codificação das expressões faciais humanas em unidades objetivas, para que essas sirvam como base na adaptação do sinalizador automático para a apresentação das emoções por meio das expressões faciais do avatar 3D.

4.1 Adaptação do Tradutor Automático

O processo de adaptação de um tradutor automático existente para a adição da análise de emoções à tradução tem como foco um tradutor automático de Português Brasileiro para glosa-Libras. Nessa adaptação o processo de análise de emoções é realizado sobre o mesmo texto em Português Brasileiro utilizado pelo tradutor automático para Libras, uma vez que a informação em si não é modificada, apenas sua forma de transmissão. Isso ocorre porque informações como sentimento e emoções não são, e nem devem ser, modificadas pela tradução automática. Além disso, existem mais recursos disponíveis voltados para a língua portuguesa do que diretamente para Libras.

As tarefas de análise de emoções e tradução automática são independentes entre si, havendo apenas dependência dos seus resultado para a construção da resposta final. Com isso, para a definição desse sistema de tradução automática de Português Brasileiro para glosa Libras adaptado, é possível a realização da análise da emoção de forma paralela à tradução automática.

De forma mais didática, na Figura 6 é apresentado um exemplo do fluxo de tradução texto-glosa adaptado para a realização da classificação da emoção do texto. A Figura ilustra, como mencionado anteriormente, que nesse serviço de tradução adaptado o mesmo texto de entrada é utilizado no tradutor automático para LS e no classificador de emoções. No final do processo a emoção identificada pelo algoritmo será então aplicada à tradução gerada pelo tradutor automático.

Ainda em relação a Figura 6, os módulos de Classificação de Emoções e de Tradução são independentes, de forma que não importa se ambos são executados no mesmo ambiente ou com as mesmas tecnologias, contanto que sejam executados paralelamente.

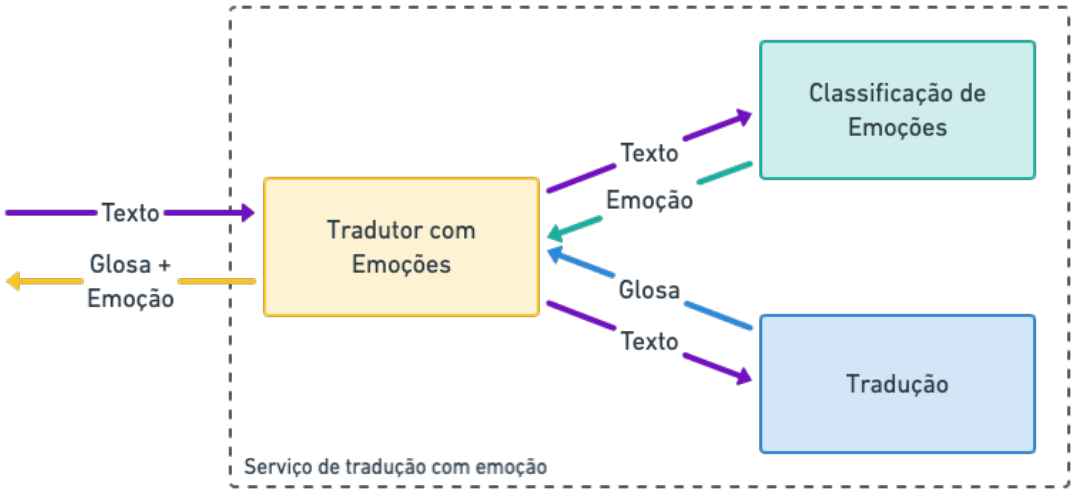

Figura 6: Exemplo do fluxo de tradução de um texto para glosa em conjunto com a análise de emoções

Para que a nova glosa com emoção seja interpretada pelo sinalizador automático do tradutor, é necessário que a emoção seja referenciada na glosa seguindo uma sintaxe. Além disso, é necessária a adaptação de um sinalizador automático para que a informação das emoções também seja expressada durante a sinalização, que para esse caso foi escolhida a abordagem de apresentação das emoções por meio das expressões faciais do avatar 3D.

Na próxima seção será explorada uma sintaxe para a adição da informação da emoção à glosa traduzida pelo tradutor automático. Já na seção subsequente, será apresentada uma forma de codificação das expressões faciais humanas e que serviram de base para adaptação das expressões faciais do avatar 3D para a apresentação das emoções.

4.2 Sintaxe de Adição da Emoção na Tradução

Em geral, em uma aplicação de tradução automática de língua oral para língua de sinais baseada na utilização de um agente virtual para sinalização, o responsável por

realizar a requisição para a tradução automática do conteúdo e de performar a tradução resultante é o sinalizador automático.

Dessa forma, para a adição da emoção à tradução, é necessária a definição de uma sintaxe simples e lógica para ser interpretada pelo sinalizador automático. Para a adição da informação da emoção à notação glosa, tal sintaxe deve possuir características como: manter a simplicidade de utilização da glosa; ter pouco ou nenhum impacto na sintaxe já existente da glosa; ser de fácil adição e remoção, para que a emoção possa ser ignorada quando não necessária; e que seja escalável, caso necessite futuras evoluções.

Nesse sentido, foi definida uma sintaxe baseada em *tags* (ou marcadores) simples, como as utilizadas nas sintaxes de linguagens de marcação como HTML e XML. Os marcadores são incorporados à glosa sempre em pares, onde a emoção entre os caracteres $<$ e $>$ denota o início da emoção daquela sequência de texto, e a emoção entre $</$ e $>$, denota o fim da emoção para aquela sequência de texto. Um exemplo de uso simples da sintaxe pode ser observada na Figura 7 (a).

Um exemplo de um uso mais complexo dessa sintaxe é apresentada na Figura 7 (b), onde são utilizada duas *tags* encadeadas, de forma que a $< EMOÇÃO\_2 >$ pode sobrepor a $< EMOÇÃO\_1 >$ durante o seu escopo, ou até mesmo as emoções podem ser apresentadas de forma conjunta, dependendo de como o interpretador dessa sintaxe seja implementado.

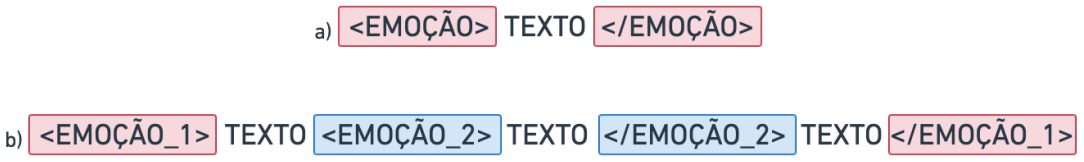

Figura 7: Exemplos de utilização da sintaxe para adição da emoção na glosa

A sintaxe definida mantém a simplicidade da glosa, tendo pouco impacto nela, além de poder ser simplesmente ignorada pelo sinalizador automático, se necessário. Ela pode ser facilmente adicionada a uma glosa, podendo ser composta de forma simples ou mais complexa, como apresentado na Figura 7. Por fim, essa sintaxe de marcadores, assim como no HTML e XML, possibilita uma futura ampliação do seu uso com ainda mais informações sobre a emoção na forma de propriedades ou metadados.

Na próxima seção é apresentada como ocorreu a codificação das expressões faciais das emoções humanas básicas que serviram como base para a definição das expressões faciais do avatar 3D.

4.3 Codificação das Expressões Faciais para a Sinalização Automática

Uma forma simples de se demonstrar as emoções durante a sinalização das traduções é utilizando a estratégia de pistas visuais, como a adotada por Lemos [31]. Porém, para essa pesquisa optou-se pela demostração das emoções da tradução por meio das expressões faciais do avatar 3D.

Para tal, é tomado como base o trabalho de Silva et al. [59], onde é apresentada uma tabela que codifica as expressões faciais presentes na Libras no formato da notação do sistema FACS (*Facial Action Coding System*), que foi primariamente definido por Ekman e Friesen [16]. Pelo FACS, qualquer expressão facial pode ser descrita em termos de unidades de ação (do inglês, *action units*) (AUs), que representam os movimentos musculares necessários para a realização de tal ação na face. Além disso, essa notação também é composta da intensidade da AU, que pode ser representada por uma letra entre A (intensidade leve) e E (intensidade máxima).

As expressões faciais da Libras tabeladas por Silva et al. [59] são dividias em: Expressões Básicas, como abir boca, juntar sobrancelhas e abrir os olhos; e Expressões Compostas, que se subdividem em Expressões Faciais Gramaticais da Sentença, como questão, negação e afirmação, e Expressões Faciais Afetivas, como a expressão de emoções básicas de alegria, tristeza, raiva, etc.

Para a definição das expressões faciais de emoções na tabela, Silva et al. [59] utilizam valores já definidos na literatura para a expressão das emoções humanas básicas, atentando-se para o fato de que as emoções básicas são amplamente utilizadas e compreendidas pela comunidade surda.

A Tabela 1 apresenta a porção da tabela definida em Silva et al. [59], referente à definição das emoções básicas em termos de AUs. Como as Expressões Faciais Afetivas são Expressões Compostas, são necessários alguns conjuntos de AUs para as descrever.

Tabela 1: Tabela de descrição das emoções básicas na notação de FACS, adaptada de Silva et al. [59].

Expressões Faciais Afetivas		
	FACS	Descrição
Emoções Básicas	AU6 + AU12	Alegria
	AU1 + AU4 + AU15	Tristeza
	AU4 + AU5 + AU7 + AU23	Raiva
	AU1 + AU2 + AU5B + AU26	Surpresa
	AU1 + AU2 + AU5 + AU20 + AU26	Medo
	AU1 + AU4 + AU5 + AU7	
	AU9 + AU15 + AU16	Desgosto

Cada AU se refere a um atuador para realizar uma determinada expressão facial. De forma resumida, *AU1* se refere ao atuador levantador de sobrancelha interna, *AU2* ao levantador de sobrancelha externa, *AU4* ao abaixador de sobrancelha, *AU5* ao levantador de pálpebra superior, *AU7* ao apertador de pálpebra, *AU9* ao enrugador de nariz, *AU12* ao puxador do canto dos lábios, *AU15* ao depressor do canto dos lábios, *AU16* ao depressor do lábio inferior, *AU20* ao esticador de lábios, *AU23* ao endurecedor de lábios e *AU26* ao atuador responsável pela queda do queixo.

Essas descrições das expressões faciais das emoções básicas serviram como base para a representação das emoções através de movimentos equivalentes nas expressões faciais do avatar 3D durante a adaptação do sinalizado automático.

A implantação e utilização dessa adaptações serão abordadas no próximo Capítulo, onde será a apresentada a construção e aprimoramento do analisador de emoções e um protótipo de adaptação do tradutor automático da Suíte VLibras utilizando os conceitos e definições apresentadas nessa seção.

5 AVALIAÇÃO EXPERIMENTAL

5.1 Desenvolvimento do Analisador de Emoções

A análise das emoções dos textos foi realizada utilizando uma rede neural treinada com uma base de dados especializada para o problema. A criação de uma arquitetura de rede neural especializada para a análise de emoções requer um grande esforço de pesquisa e experimentos. Assim, uma alternativa para isso é a utilização de uma arquitetura para um problema equivalente, onde só é preciso adaptar o que for necessário.

Na próxima seção será descrita de forma mais detalhada como é composta a rede neural para a análise de emoções em textos em Português Brasileiro utilizada nesse estudo, partindo de um método voltado para a classificação de textos.

5.1.1 Desenvolvimento do Algoritmo para Análise de Emoções

Como explanado na Seção 2.3.1, a análise de emoções de textos pode ser visualizada como um problema de classificação textual. Dito isso, optou-se pela utilização do método MultiFiT (*Multi-lingual Fine-Tuning*), de Eisenschlos et al. [15], visto que sua criação foi pensada para a utilização de classificação de textos em qualquer idioma. Ele é uma solução mais robusta que o método ULMFiT, no qual foi inspirado. Ambos os métodos atingiram excelentes resultados em seus testes e apesar do MultiFiT não se manter como estado da arte, ele ainda é uma solução relevante e mais leve que as demais, além de possuir mais recursos de fácil acesso para o Português.

O MultiFiT utiliza a estratégia de *transfer learning* na sua construção. A ideia geral dessa estratégia é utilizar múltiplas redes neurais de modo que se reduza a complexidade do problema a cada rede utilizada. Na definição do MultiFiT para a tarefa de classificação, ele é composto de duas NNs, onde a primeira é uma rede para *Language Model*, que receberá os dados de entrada a serem utilizados na especialização da rede, cujo o resultado será utilizado na segunda rede neural responsável pela classificação. Essa segunda rede neural usará o que foi aprendida pela primeira para realizar a classificação dos textos.

A Figura 8 apresenta uma visão geral de todo o processo do MultiFiT: o primeiro passo (1) é o treinamento de uma rede LM bidirecional com uma base de dados de propósito geral da língua alvo, na qual será realizado o procedimento de *fine-tuning* (ou afunilamento); (2) para o contexto trabalhado, será feito uma especialização da rede neural na base de dados com a qual se está trabalhando, que é a mesma base de dados que será utilizada no processo de classificação, porém sem as classes, apenas os textos de entrada; a próxima etapa (3) é utilizar o conhecimento adquirido pela rede LM, ou seja, os pesos da rede, e o utilizar no processo de afunilamento da rede de classificação, usando a mesma base que anteriormente, mas agora com as classes dos textos.

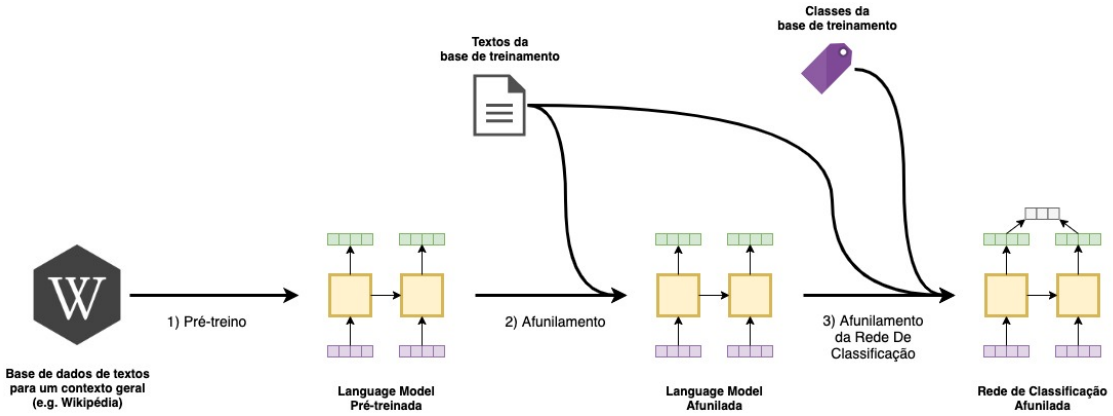

Figura 8: Etapas do MultiFiT

Como indicado no trabalho de Eisenschlos et al. [15], na implementação serão utilizadas redes QRNN (Quasi-Recurrent Neural Networks) [5], um tipo de rede convolucional que pode ser paralelizada e que utiliza funções de *pooling* recorrentes. Essas redes superam a performance das LSTMs tradicionais e são computacionalmente mais rápidas e leves. Tanto a rede de *Language Model*, quanto de classificação utilizam 4 camadas de redes QRNN com uma dimensionalidade de 1.550 camadas ocultas.

Na próxima seção será descrito o processo de aquisição e preparação da base de dados que foi utilizada para treinamento e teste da rede de classificação de emoções.

5.1.2 Definição da Base de Dados

Um requisito comum às estratégias que se utilizam de NNs é a necessidade de uma base de dados abrangente o suficiente para a solução do problema. Existem pouquíssimas bases de dados em Português Brasileiro voltadas para a classificação de emoções, sendo mais fácil encontrar bases de dados voltas para a classificação de sentimentos, ou seja, que classificam os textos puramente em sentimento positivo, negativo e por vezes o sentimento neutro.

No trabalho de Dosciatti, Ferreira e Paraiso [14], um algoritmo de SVM é treinado para a classificação de emoções de textos em Português. O conjunto de dados é composto de manchetes do site *www.globo.com*, de diversas categorias, contando com um total de 1.750 textos, sendo 250 para cada uma das 6 emoções abordadas (alegria, desgosto, medo, raiva, surpresa e tristeza), escolhidas com base no modelo de Plutchik, e 250 para a classe "neutro". O processo de rotulação dos textos foi realizado por duas pessoas: uma especialista em linguística e outra em linguística computacional. A base de dados desse trabalho pode ser adquirida através de uma solicitação no site do projeto [3].

[3]http://www.ppgia.pucpr.br/ paraiso/mineracaodeemocoes/recursos/g1$_v$2.*php*

A quantidade de textos rotulados por esses autores [14] é pequena em comparação a outros trabalhos que realizam a tarefa de classificação de emoções em texto. Diante disso, foi realizado um processo de *data augmentation* dessa base de dados.

A realização de *data augmentation* é muito comum em treinamento de redes neurais que utilizam imagens como dados de treinamento. Em suma, essa técnica consiste em processos que modificam um dado existente na base de dados de treinamento para que ele se torne um novo dado também relevante para a base de treinamento. No contexto de imagens, são comuns processos como rotação, espelhamento, mudança dos canais de cores etc. Por outro lado, no contexto de dados textuais, modificações semelhantes não podem ser realizadas, pois mudariam o dado a ponto de os deixar errados, de modo a atrapalhar o aprendizado do algoritmo.

A utilização de *data augmentation* para a ampliação da base de dados será explorado na próxima Seção, onde será apresentado o planejamento dos experimentos para realizar o aprimoramento do analisador de emoções com rede neural.

5.1.3 Planejamento Experimental

Nessa Seção foram realizados alguns experimentos computacionais para o aprimoramento da rede neural para a análise de emoções de textos em Português Brasileiro, visando encontrar a melhor configuração de parâmetros do algoritmo de *data augmentation* para a definição da base de treinamento da rede neural.

A heurística mais comum para realizar *data augmentation* de textos é a substituição de sinônimos. O procedimento escolhido para o presente estudo replica a estratégia definida no trabalho de Mosolova, Fomin e Bondarenko [39], que consiste em escolher as palavras no texto que podem possuir sinônimos, a partir da identificação das classes gramaticais dessas palavras, como por exemplo verbos, substantivos e adjetivos, já as palavras das demais classes gramaticais permanecem intactas. Nesse sentido, após as palavras de interesse serem identificadas, é necessário listar os sinônimos delas, para isso, utiliza-se a Wordnet[37], que é um conjunto de dados que contém informações e relacionamento das palavras da língua inglesa, para qual foi originalmente desenvolvida, mas que depois foi implementada em outras línguas, incluindo o Português Brasileiro [44].

A estratégia definida no trabalho de Mosolova, Fomin e Bondarenko [39] possui dois parâmetros para controlar a realização do processo de *augmentation* dos textos, são eles: porcentagem de palavras substituídas por iteração no texto (PctWords) e quantidade máxima de novos textos gerados por texto de origem (MaxNew).

O primeiro parâmetro (porcentagem de palavras substituídas no texto por iteração) diz respeito à porcentagem de palavras válidas para o processo de substituição que poderão ser substituídas num mesmo texto a cada iteração. Assim, a cada ciclo de *augmentation*

em um texto da base de dados, apenas uma porcentagem definida das palavras candidatas para substituição serão efetivamente substituídas, tendo seus sinônimos escolhidos aleatoriamente. Por exemplo, se um texto possui 10 palavras elegíveis para substituição e esse parâmetro definiu que apenas 20% serão substituídas, então apenas duas dessas dez palavras serão escolhidas aleatoriamente para serem substituídas. Assim, na próxima iteração outras 2 palavras serão escolhidas aleatoriamente. Caso os sinônimos escolhidos não gerem novos textos, escolhe-se novos sinônimos aleatoriamente. Se mesmo assim não for possível a geração de novos textos, novas palavras serão escolhidas.

O segundo parâmetro (quantidade máxima de novos textos gerados por texto de origem) serve como uma condição de parada do procedimento anterior. Uma forma do procedimento anterior parar é caso não seja possível a geração de novos textos, ou seja, qualquer que sejam os sinônimos escolhidos, não é possível gerar um texto distinto dos demais. Esse segundo parâmetro limita superiormente a quantidade de textos gerados.

Como todo esse processo pode gerar uma quantidade diferente de novos textos para cada classe, propõem-se também a adição de um novo parâmetro (BALANC). Esse novo parâmetro consiste em uma condição binária definindo se a nova base de dados gerada deverá ser balanceada ou não, ou seja, se todas as classes devem conter a mesma quantidade de textos ou não. Caso positivo, a quantidade de textos de cada classe será limitada pela classe com menos ocorrências.

Dessa foram, define-se um Planejamento de Experimentos Fatorial 2^k, onde são definidos k fatores de testes que podem influenciar o resultado final. Para cada fator, são definidos dois valores (ou níveis). A partir disso, são planejados 2^k experimentos distintos onde se variam os níveis dos fatores.

Na Tabela 2 são apresentados os fatores definidos para os experimentos, bem como os seus níveis. Para o fator PctWords são definidos os valores 25% e 75%, sendo o menor nível originado de Mosolova, Fomin e Bondarenko [39] e o maior nível foi definido como um valor de controle. Para o fator MaxNew são definidos os valores 6 e 10, também seguindo os mesmo critérios do anterior. Já o fator BALANC possui um valor binário, sendo esse o balanceamento ou não da base de dados.

Tabela 2: Níveis dos Fatores

Fator	-1	1
PctWords	25%	75%
MaxNew	6	10
BALANC	Não	Sim

A Tabela 3 apresenta a tabela de experimentos a serem realizadas.

Tabela 3: Matriz de Planejamento

Nº do Experimento	PctWords	MaxNew	BALANC
1	-1	-1	-1
2	-1	-1	1
3	-1	1	-1
4	-1	1	1
5	1	-1	-1
6	1	-1	1
7	1	1	-1
8	1	1	1

5.1.3.1 Métricas de Interesse

Os resultados serão avaliados tendo como métrica de interesse a acurácia de acerto das emoções e o *F1-score*, que leva em consideração a precisão e *recall* das predições das redes.

Para a análise dos resultados foi realizada uma Análise da Variância (ANOVA), visando estudar a influência e significância dos fatores nas métricas computacionais geradas. Outro passo da análise foi o cálculo do coeficiente de determinação (R^2) para verificar a relevância do modelo sobre os resultados alcançados.

5.1.3.2 Configuração do Ambiente

A arquitetura MultiFiT foi implementada utilizando o *framework* fastai[4], uma abstração em auto-nível do *framework* de *machine learning* PyTorch[5]. Como indicado no trabalho de Eisenschlos et al. [15], tanto na implementação da rede de *Language Model*, quanto na rede de classificação foram utilizadas 4 camadas de redes QRNN com uma dimensionalidade de 1.550 camadas ocultas.

A rede *Language Model* bidirecional utilizada[6] foi pré-treinada utilizando a base de dados de textos em Português do Wikidata[7], que reúne textos de artigos do Wikipédia em diversas línguas, limitando-se o vocabulário a 15.000 *tokens* e utilizando um vetor *embedding* de tamanho 400.

A implementação da heurística de *data augmentation* se deu na linguagem Python, utilizando as bibliotecas SpaCy[8] e NLTK [33], para a identificação das classes gramaticais e como interface para acessar os sinônimos das palavras na Wordnet em Português.

[4]https://github.com/fastai/fastai/
[5]https://pytorch.org/
[6]Disponível em https://github.com/piegu/language-models
[7]https://www.wikidata.org/
[8]https://spacy.io/

A partir da base de dados original de Dosciatti, Ferreira e Paraíso [14], foram separados 10% dos exemplos para a criação de uma base de dados de teste e 90% para *data augmentation*. Ambas as bases de dados, para teste e para *data augmentation*, iniciaram-se balanceadas.

Todos os experimentos foram executados no ambiente do Google Colab[9].

Os resultados referentes aos experimentos planejados serão apresentados no Capítulo 6. Já na próxima Seção será apresentada a adaptação da ferramenta de tradução VLibras através da aplicação dos processos de adaptação descritos no Capítulo anterior.

5.2 Protótipo de Adaptação do Tradutor Automático da Suíte VLibras

Objetivando aplicar a estratégia de análise de emoções em conjunto com a adaptação de uma ferramenta de tradução automática de língua oral para língua de sinais, nesse capítulo será apresentado o protótipo da adaptação da Suíte VLibras, que realiza a tradução automática de conteúdos multimídia em Português Brasileiro para Libras, a fim de possibilitar a interpretação de emoções atribuídas à tradução.

Os fundamentos necessários para a adaptação de um ferramenta de tradução automática de LO para LS já foram apresentados na Seção 5.2. A seguir será descrito como os conceitos e definições foram aplicados na adaptação do serviço em nuvem para a tradução automática de textos em Português Brasileiro para glosa-Libras, bem como na versão *desktop* do sinalizador automático, o qual faz uso de um avatar 3D, onde ambos compõem a Suíte VLibras [2].

5.2.1 Adaptação do Tradutor Automático da Suíte VLibras

Seguindo a arquitetura genérica apresentada na Figura 6 da Seção 4.1, foi implementado um serviço HTTP simples, utilizando o modelo de classificação de emoções que obteve o melhor resultado nos experimentos da Seção 5.1.3, e o serviço em nuvem de tradução de texto da Suíte VLibras, o qual é acessado por meio de requisição a uma API.

De forma mais específica, o módulo responsável pela classificação das emoções dos textos estará sendo executado no mesmo ambiente do servidor construído, enquanto que o módulo de tradução realizará as requisições ao serviço de tradução da Suíte VLibras, que está sendo executado em um ambiente externo, o qual é o mesmo serviço utilizado pelas demais ferramentas da Suíte VLibras.

Além disso, os processos de classificação de emoção e tradução são executados de forma paralela. O procedimento de classificação é realizado em dois subprocessos

[9]https://colab.research.google.com/

paralelos, um para a classificação do texto no sentido direto (*forward*) e outro para a classificação no sentido inverso (*backward*). No fim é realizada uma média dos pesos resultantes de cada subprocesso para a classificação final.

Uma vez finalizada a análise de emoção do texto e a tradução, a emoção identificada é adicionada à glosa resultante da tradução, utilizando a sintaxe definida na Seção 4.2. Para esse protótipo é realizada a classificação de apenas 1 sentimento por texto.

A Figura 9 apresenta um diagrama de classes do servidor implementado. Esse serviço de tradução adaptado foi criado utilizando a linguagem Python e o *micro-framework Flask*, para prover um API de acesso. Além disso, foram utilizadas bibliotecas para auxiliar no paralelismo das operações e o *framework* fastai para o uso dos modelos treinados.

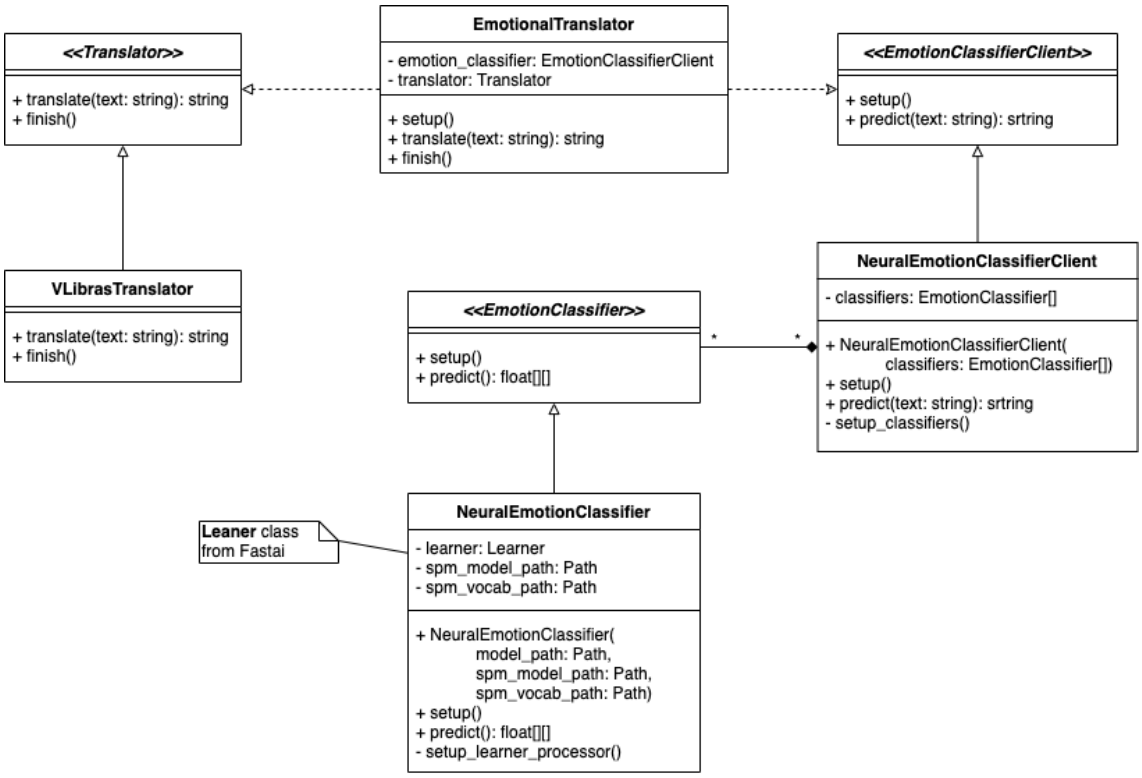

Figura 9: Digrama de classes da implementação do serviço de tradução adaptado

O papel do módulo de classificação de emoções é exercido principalmente pela classe $NeuralEmotionClassifierClient$, que utiliza os classificadores de emoções baseados em redes neurais para identificar a emoção de um dado texto. Esses classificadores são implementados utilizando a classe $NeuralEmotionClassifier$, que é responsável por carregar e encapsular o modelo de rede neural treinado anteriormente, ou seja, os modelos *forward* e *backward*

A classe $VLibrasTranslator$ exerce o papel do módulo de tradução, sendo responsável por fazer as requisições de tradução para o serviço em nuvem da Suíte VLibras.

Por fim, a classe *EmotionalTranslator* é responsável por realizar a tradução de um texto em Português Brasileiro para glosa-Libras, com a adição da emoção identificada para esse mesmo texto, utilizando-se dos módulos de classificação de emoções e tradução.

A implementação desse sistema foi realizada de forma a separar bem os módulos de classificação de emoções e de tradução através das interfaces, de tal modo que a modificação e troca desses módulos possa ser realizada de forma simples e com mínimo impacto na arquitetura.

Na próxima seção será descrita a adaptação do sinalizador automático da versão *desktop* da Suíte VLibras para a apresentação das emoções presentes na glosa.

5.2.2 Adaptação do Sinalizador Automático do VLibras

Nesta segunda etapa da adaptação do sistema de tradução VLibras para a interpretação de emoções em textos foi necessária uma alteração no código-fonte da versão *desktop* da ferramenta de sinalização automática, fornecido pelo Lavid (Laboratório de Vídeo Digital)[10], o qual é o desenvolvedor do projeto.

Para a viabilidade da construção do protótipo de adaptação do sinalizador automático, foram observadas quatro ações principais que deveriam ser realizadas, sendo essas:

- Utilizar o serviço de tradução de textos em Português Brasileiro para glosa-Libras com emoção;
- Interpretar a emoção adicionada à glosa através da sintaxe definida na Seção 4.2;
- Demonstrar as emoções por meio das expressões faciais do avatar;
- Alterar a interface do usuário para que a apresentação da emoção seja ativada e desativada.

A primeira ação, referente à utilização do serviço de tradução com análise de emoções, foi realizada por meio da alteração do endereço da API responsável por realizar a tradução em nuvem utilizada pelo sinalizador, para que ele utilizasse o serviço de tradução adaptado. Nesse protótipo o serviço de tradução automática foi executado localmente e utilizando o mesmo formato de requisição do endereço da API original de tradução, atuando assim como um *middleware* entre a ferramenta e a API de tradução original da Suíte VLibras.

[10]https://lavid.ufpb.br/

A segunda ação, que se refere a interpretação das emoções na glosa, foi abordada modificando um pouco como a glosa resultante da tradução era interpretada pelo sinalizador. Para a ferramenta, cada palavra da glosa se refere a uma animação, chamada de *bundle*, sendo esse que define como o avatar 3D deve performar em tela. Além disso, as animações subsequentes são interpoladas pela *engine* Unity[11], na qual a aplicação é construída. Por fim, durante a sinalização, uma legenda na parte inferior indica qual sinal está sendo performado no momento. Ressalta-se que essa legenda também é obtida através da glosa.

O desafio nesse cenário é identificar quando a emoção da glosa deve ser performada ou não. Além disso, outra questão é que essa informação não deve aparecer na legenda que acompanha a sinalização. Todavia, frente a esses problemas, da forma como a sintaxe para a adição das emoções na glosa foi definida na Seção 4.2, existem caracteres especiais que definem o início e fim de uma dada emoção no texto. Com isso, é fácil identificar quando uma palavra na glosa se refere a uma animação ou emoção. Quando identificada que a palavra na glosa se refere ao inicio ou fim da emoção do texto, essa palavra não é buscada na lista de animações e nem é apresentada na legenda, caso contrário, a animação é performada normalmente.

Como também é definida na sintaxe da Seção 4.2, é possível a ocorrência de emoções encadeadas, apesar disso não ser abordado na implementação apresentada na Seção 5.2.1. Para tal, foi definido que as emoções encadeadas irão sobrepor a execução da emoção atual. As emoções identificadas são postas numa lista e são removidas quando chegam ao seu fim, ou a sinalização do texto é finalizada.

Um exemplo disso pode ser observado na Figura 10, onde em um primeiro momento a $< EMOÇÃO\_1 >$ é performada, mas num dado instante do texto, a $< EMOÇÃO\_2 >$ aparece, sobrepondo a primeira emoção até que $< /EMOÇÃO\_2 >$ aparece, que é quando $< EMOÇÃO\_1 >$ voltará a ser performada, até $< /EMOÇÃO\_1 >$ aparecer.

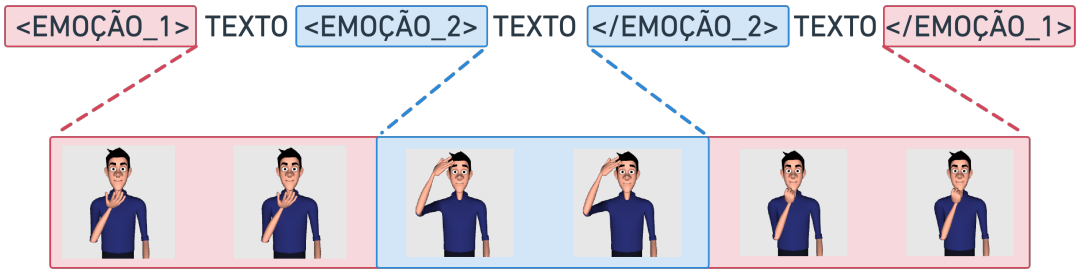

Figura 10: Exemplo da interpretação da glosa com emoção na adaptação do sinalizador automático do VLibras

Para realizar a terceira ação, referente à demonstração das emoções por meio de

[11]https://unity.com/

expressões faciais do avatar 3D, assim como na ação anterior, também foi necessária uma modificação no código-fonte da ferramenta. As expressões faciais já estão presentes no avatar, as quais aparecem em alguns sinais específicos. Como dito na Seção 2.2, expressão facial é um dos articuladores utilizados na construção gramatical de um sinal em uma língua de sinais.

Na ferramenta, as expressões faciais do avatar estão presentes nos *bundles* de animações dos sinais. Uma solução pouco viável seria a recriação dos *bundles* de cada sinal adicionando um pouco mais de expressão facial para denotar as emoções específicas, dado que existem milhares de sinais e esse número seria multiplicado pela quantidade de emoções abordadas, além de ser um trabalho manual.

Como dito anteriormente, a ferramenta de sinalização do VLibras foi construída na plataforma Unity, a qual possui suporte à ferramenta Blender[12], sendo essa a ferramenta em que o avatar foi modelado e que também é utilizada pelos animadores do projeto para criar os *bundles* das animações.

Para facilitar o processo de animação das expressões faciais, o modelo do avatar 3D já contém algumas animações pré-definidas para movimentar partes específicas do rosto, as quais podem ser acessadas e modificadas programaticamente pelo Unity. Essas animações pré-definidas são responsáveis por realizar movimentos como: sorriso, fechar sobrancelhas, abrir sobrancelhas, abaixar sobrancelhas, subir sobrancelhas, franzir sobrancelhas, subir lábio superior, contrair lábios, abaixar cantos da boca, subir cantos da boca, inflar bochechas, fazer bico, e contrair bochechas.

Essas animações pré-definidas possuem um valor em ponto flutuante que representa o seu progresso, isto é, 0.0 representa o início da animação e 100.0 o fim. Utilizando o Unity é possível modificar esses valores a qualquer momento, o que será feito quando identificada alguma emoção na glosa.

Pode-se verificar que essas animações de expressões faciais pré-definidas se assemelham às AUs descritas na Seção 4.3. Apesar do sinalizador não possuir animações equivalentes a todas as AUs, elas ainda podem ser tomadas como base para a definição das expressões faciais das emoções no avatar 3D, com algumas adaptações quando necessário.

A partir dessas animações pré-definidas, foi possível a definição de todas as emoções abordadas na pesquisa, apesar de algumas adaptações terem sido necessárias para alcançar resultados semelhantes às descrições da Tabela 1 da Seção 4.3. Vale considerar que a expressão padrão do avatar na maioria dos sinais é considera neutra, o que também compreende uma das emoções abrangidas pelo analisador de emoções.

[12]https://www.blender.org/

Para expressar Alegria foram modificadas os valores das animações de sorriso e subir os cantos da boca. Almejando expressar Tristeza, foram modificados os valores das animações de abrir sobrancelhas, abaixar sobrancelhas, abaixar cantos da boca e contrair lábio. Visando expressar Raiva, foram modificados os valores das animações de franzir sobrancelhas, fechar sobrancelhas, subir sobrancelhas, fazer bico e contrair os lábios. Objetivando expressar Medo, foram modificadas as animações de abrir sobrancelhas, subir sobrancelhas, contrair bochechas, e também foi utilizada a animação de sorriso, que possibilita movimentar um pouco o lábio inferior, e os demais movimentos do sorriso foram corrigidos com as animações de subir lábio superior e abaixar cantos da boca. Visando expressar Desgosto, foram modificados os valores das animações de abaixar cantos da boca, contrair lábios, e também a animação de franzir as sobrancelhas para poder enrugar um pouco o nariz; essa animação foi corrigida com a animação de abrir as sobrancelhas. Por fim, intentando expressar Surpresa, foram modificados os valores das animações de abrir sobrancelhas, subir sobrancelhas e subir lábio superior.

Por fim, para realizar a quarta ação, referente à alteração da interface do usuário para que ele possa ativar e desativar a apresentação das emoções, foi adicionado um botão à barra inferior, juntamente aos demais controles de reprodução da ferramenta.

No próximo Capítulo serão abordadas os resultados referentes ao experimentos planejados na Seção 5.1.3 e a da adaptação da ferramenta de tradução automática da Suíte VLibras apresentada na Seção 5.2.

6 ANÁLISE DOS RESULTADOS

6.1 Resultados dos Experimentos

Conforme definido na Tabela 3 da Seção 5.1.3, os experimentos planejados visaram a avaliação dos fatores: porcentagem de palavras substituídas no texto por iteração (PctWords), quantidade máxima de novos textos por texto de origem (MaxNew) e balanceamento da base de dados (BALANC), objetivando a realização de *data augmentation* da base de dados de Dosciatti, Ferreira e Paraíso [14].

A Tabela 4 traz os resultados das métrica de acurácia e *F1-score* dos 8 experimentos, sendo esses executados em duas baterias de testes ($R1$ e $R2$) utilizando o mesmo ambiente. Nos resultados apresentados é possível observar que o Experimento 8 (PctWord de 75%, MaxWord de 10 e balanceamento ativado) da primeira bateria ($R1$) obteve os melhores resultados, tanto em acurácia (94,29%), quanto no *F1-score* (94,28%), sendo esses resultados consideravelmente superiores aos resultados obtidos por Dosciatti, Ferreira e Paraíso [14], que foram de 60,7% de acurácia e 60% de *F1-score*, utilizando uma estratégia de SVM.

Tabela 4: Resultados de Acurácia e *F1-score* dos Experimentos

Nº do Experimento	R1		R2	
	Acurácia	F1-score	Acurácia	F1-score
1	68,00%	67,77%	63,43%	63,13%
2	93,14%	93,15%	90,29%	90,28%
3	67,43%	67,00%	62,86%	62,42%
4	93,14%	93,20%	**91,43%**	**91,48%**
5	66,29%	65,45%	60,00%	58,71%
6	93,71%	93,74%	90,86%	90,94%
7	68,00%	67,65%	60,57%	60,83%
8	**94,29%**	**94,28%**	90,86%	90,85%

Na Tabela 5 são apresentados, de forma detalhada, os resultados das métricas de acurácia e *F1-score* de cada emoção no Experimento 8 de $R1$ que aqui foram realizados, e dos resultados alcançados no trabalho de Dosciatti, Ferreira e Paraíso [14], havendo assim uma comparação. Nota-se que, no experimento aqui desenvolvido, para as emoções "raiva"e "surpresa", houve uma acurácia de 100%, mas um *F1-score* de 98,04% em ambos os casos. Dessa forma, é possível visualizar a superioridade dos resultados da proposta em relação ao trabalho de Dosciatti, Ferreira e Paraíso [14], em todas as emoções trabalhadas, uma vez que foram obtidos valores de *F1-score* superiores a 90% em todos os casos, havendo assim uma alta taxa de acerto nas classificações.

Tabela 5: Comparação dos Resultados do Experimento 8 da Primeira Bateria (Exp8 (R1)) de Experimentos com os Resultados de Dosciatti, Ferreira e Paraíso [14] (DFP)

Emoção	Acurácia		F1-score	
	Exp8 (R1)	DFP	Exp8 (R1)	DFP
Alegria	92%	45%	93,88%	46%
Desgosto	88%	39%	91,67%	40%
Medo	92%	81%	93,88%	76%
Neutro	96%	50%	90,57%	51%
Raiva	100%	75%	98,04%	75%
Surpresa	100%	81%	98,04%	78%
Tristeza	92%	54%	93,88%	54%
Média	94,29%	61%	94,28%	60%

Em relação à segunda rodada de experimentos (R2), o Experimentos 4 (PctWord de 25%, MaxWord de 10 e balanceamento ativado) obteve uma acurácia de 91,43% e *F1-score* de 91,48%, sendo o melhor dessa bateria de experimentos, mas ainda assim, não tão bom quanto o Experimento 8 em R1.

Além disso, é aparente a diferença dos resultados dos experimentos de números pares em relação aos de números ímpares. Nos experimentos pares, o único fator que se mantém com o mesmo valor é a utilização de uma base de dados balanceada, ao ponto que os experimentos ímpares não possuem uma base de dados balanceada. Isso pode apontar que o fator de balanceamento da base de dados pode ter uma significativa influência nos resultados obtidos. De fato, essa suposição pode ser comprovada através de uma Análise de Variância (ANOVA) dos resultados, a qual é apresentada nas Tabela 6, referente à acurácia, e Tabela 7, referente ao F1-score.

Tabela 6: ANOVA da Métrica de Acurácia dos Experimentos

Origem	SS Parcial	*df*	*MS*	F	Prob >F
Modelo	0,30648	7	0,04378	41,82	0,0000
PctWord	0,00165	1	0,00016	0,16	0,7015
MaxNew	0,00005	1	0,00005	0,05	0,8308
PctWor*MaxNew	0,00005	1	0,00005	0,05	0,8308
BALANC	0,30564	1	0,30564	291,95	**0,000**
PctWord*BALANC	0,00046	1	0,00046	0,44	0,5264
MaxNew*BALANC	$2,0408e^{-6}$	1	$2,0408e^{-6}$	0,00	0,9659
PctWord*MaxNew*BALANC	0,0001	1	0,0001	0,01	0,7652
Residual	0,00837	8	0,00105		
Total	0,31485	15	0,02099		

Tabela 7: ANOVA da Métrica *F1-score* dos Experimentos

Origem	SS Parcial	*df*	*MS*	F	Prob >F
Modelo	0,31763	7	0,04537	43,94	0,0000
PctWord	0,00223	1	0,00022	0,22	0,6545
MaxNew	0,00013	1	0,00013	0,12	0,7333
PctWor*MaxNew	0,00016	1	0,00016	0,15	0,7069
BALANC	0,31629	1	0,31629	306,32	**0,000**
PctWord*BALANC	0,000548	1	0,00055	0,53	0,4870
MaxNew*BALANC	$8,2199e^{-6}$	1	$8,2199e^{-6}$	0,01	0,9311
PctWord*MaxNew*BALANC	0,00027	1	0,00027	0,26	0,6218
Residual	0,00826	8	0,00103		
Total	0,32589	15	0,02172		

Em ambas as tabelas de Análise de Variância, Tabela 6 e Tabela 7, é possível verificar que o fator BALANC é o que possui uma maior significância estatística sobre os resultados apresentados, mostrando que esse é o fator que possui uma influência considerável nos resultados. Além disso, por meio das análises, é possível verificar que em relação aos demais fatores e suas interações, esses não possuem significância estatística sobre os resultados, de modo que a mudança de seus valores não exercem influência significativa nos resultados apresentados, nem mesmo as interações com o fator BALANC, o qual possui maior significância estatística isoladamente.

Em relação à análise sobre as acurácias, o modelo obteve um R^2 de *0,9734* e o R^2 ajustado de *0,9501*. Já no que tange a análise sobre os *F1-score*, o modelo obteve um R^2 de *0,9747* e o R^2 ajustado de *0,9525*. Ambos os casos apontam que os modelos se adéquam aos dados apresentados.

6.2 Resultados da Adaptação da Suíte VLibras

A falta de animações específicas para a definição das expressões faciais do avatar 3D tal qual as utilizadas nas faces humanas prejudicou a apresentação de algumas emoções. Por exemplo, a ausência do acesso a movimentos como o abrir da boca e controle da pálpebra, frente às animações pré-definidas do avatar, prejudicou a definição de algumas expressões como medo, surpresa e desgosto.

Um exemplo da performance das expressões faciais da emoções no avatar 3D da Suíte VLibras é apresenta na Figura 11, que traz *frames* da sinalização da palavra *ANDAR* nas emoções neutra (sinalização padrão da ferramenta), alegria, tristeza, raiva, medo, desgosto e surpresa.

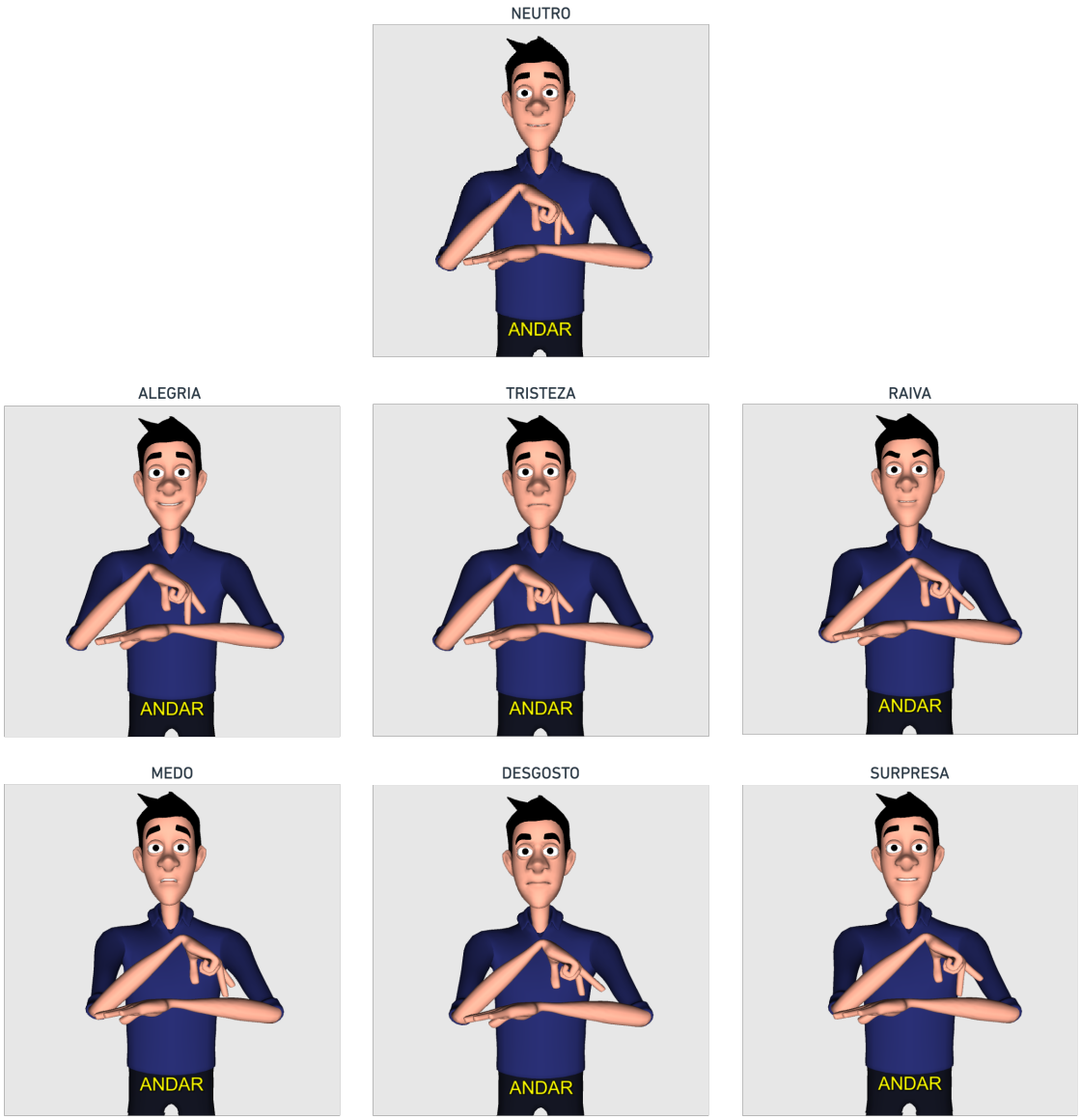

Figura 11: Exemplo da sinalização da palavra ANDAR nas emoções neutra, alegria, tristeza, raiva, medo, desgosto e surpresa.

Vale ressaltar que os testes foram realizados apenas com o avatar masculino, visto que, apesar da ferramenta conter uma avatar feminino, ele não apresenta as mesma animações pré-definidas para as expressões faciais, fazendo com que elas não sejam tratadas da mesma forma, fato que inviabilizaria a aplicação nesse avatar em específico, caso fossem utilizados os mesmos processos aqui apresentados.

A modificação da interface do usuário para o controle da apresentação da emoção no avatar 3D está representada na Figura 12. O botão do *smile* no final da barra de controles na parte inferior do *player* controla a execução das emoções presentes na tradução. A qualquer momento o usuário pode clicar no botão que quando aceso indica que as emoções serão performadas, e quando apagado, as emoções não serão apresentadas.

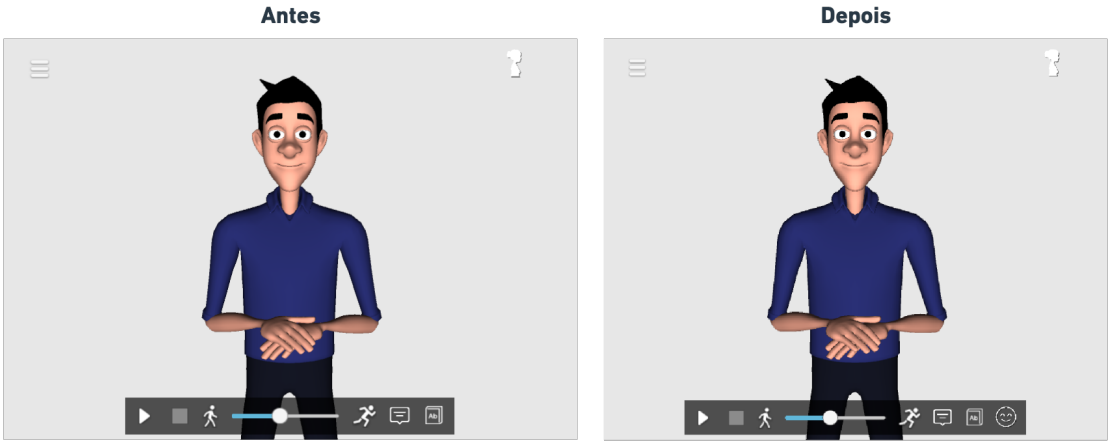

Figura 12: Comparação do Antes e Depois da interface do usuário no sinalizador automático do VLibras

Um vídeo demonstração do protótipo da adaptação da ferramenta de sinalização automática da Suíte VLibras pode ser acessado através desse link (`https://youtu.be/Kvsr4l6Yk-U`). O vídeo apresenta a sinalização de alguma frases presentes no conjunto de teste dos experimentos da Seção 5.1.3.

No próximo Capítulo serão abordadas as conclusões referentes ao trabalho apresentado, além de expor melhorias que podem ser realizadas em trabalhos futuros.

7 CONCLUSÕES E TRABALHOS FUTUROS

Em um cenário onde uma parcela significativa da população, sendo essa a população de pessoas surdas, vive à margem do acesso à ampla comunicação, visto que a maioria das informações disponíveis são voltadas prioritariamente para pessoas ouvintes, esse trabalho surge como uma maneira de tentar auxiliar no aprimoramento de técnicas computacionais capazes de estreitar a comunicação entre pessoas ouvintes e não ouvintes.

Diante disso, esse trabalho teve por objetivo explorar a estratégia de análise de emoções em textos empregando aprendizagem de máquina para a utilização em tradutores automáticos de Português Brasileiro para Libras, visando assim reduzir críticas existentes por parte da comunidade surda, a qual afirma que tradutores automáticos são benéficos, porém, a falta de humanidade e de expressão de emoções são aspectos que atrapalham a experiência dos usuários com esses tradutores.

Para atingir o seu objetivo geral, o estudo apresentou três objetivos específicos, sendo esses: explorar a construção de um sistema baseado em aprendizagem de máquina para atribuição de emoções em texto utilizando bases de dados existentes; realizar experimentos para a ampliação da base de dados de treinamento do algoritmo de aprendizagem de máquina utilizando *data augmentation*; e utilizar o sistema desenvolvido em conjunto com a adaptação de uma ferramenta de tradução automática de língua oral para língua de sinais.

O primeiro dos objetivos específicos foi alcançado por meio da utilização do método MultiFit, que utiliza a estratégia de *transfer learning* para treinar duas redes neurais responsáveis por realizar classificação de textos. Como foi explorado no decorrer do estudo, o problema de análise de emoções também pode ser visto como um problema de classificação textual. Para o treinamento da rede neural, foi utilizada a base de dados originada no trabalho de Dosciatti, Ferreira e Paraiso [14].

O segundo objetivo específico foi alcançado por meio de experimentos utilizando a heurística de *data augmentation* definida por Mosolova, Fomin e Bondarenko [39], a fim de contornar a baixa quantidade de exemplos para treinamento da base de dados de Dosciatti, Ferreira e Paraiso [14].

A definição dos fatores a serem experimentados foi realizada por meio dos parâmetros de *data augmentation* definidos no trabalho de Mosolova, Fomin e Bondarenko [39], a saber: porcentagem de palavras substituídas no texto por iteração e quantidade máxima de novos textos gerados por texto de origem. Além da adição de um novo parâmetro referente ao balanceamento da base da dados resultante do processo de *data augmentation*.

A partir dos fatores definidos, foram realizados os experimentos utilizando o Planejamento de Experimentos $2k$. Os experimentos apontaram que a utilização de uma

base de dados com *data augmentation* balanceada foi um fator significante para obter os melhores resultados, sendo esses uma acurácia de 94,29% e um *F1-score* de 94,28%, superando os resultados do trabalho que originou a base de dados, fato que mostra que a estratégia aqui abordada traz uma influência positiva para uma melhora na tarefa de análise de emoções de textos em Português Brasileiro.

O terceiro objetivo específico foi alcançado por meio da adaptação do tradutor e sinalizador automático da Suíte VLibras para a tradução de conteúdo multimídia de Português Brasileiro para Libras. A adaptação do tradutor de se deu por meio da construção de um serviço de tradução automática que utilizou o modelo de rede neural treinado para a análise de emoções de texto em Português Brasileiro e o serviço de tradução automática em nuvem da Suíte VLibras. Para isso também foi definida um sintaxe para a adição da informação de emoção na glosa-Libras.

A adaptação do sinalizador automático foi feita por meio de modificações no código-fonte da versão *desktop* do sinalizador automático da Suíte VLibras. A adaptação do sinalizador utiliza o serviço de tradução automática adaptado, interpreta a glosa com a informação da emoção e apresenta a sinalização da tradução com a emoção expressa por meio de expressões faciais no avatar 3D.

Apesar do analisador de emoções apenas tratar um pequeno contexto textual, onde obteve um ótimo resultado, ele ainda sim representa um ponto de partida interessante para a aplicação da análise de emoções de textos em Português Brasileiro em conjunto com *data augmentation*.

Da mesma forma, a adaptação do tradutor automático de textos em LO para LS, realizada a partir da arquitetura genérica apresentada e da sintaxe de construção da glosa com emoções, serve como um ponto inicial para as inúmeras possibilidades de implementações de ferramentas de tradução adaptadas. A adaptação da ferramenta de tradução automática da Suíte VLibras mostrou como essa adaptação pode ser feita com pouca interferência no tradutor automático original, mas ainda sim necessitando de uma modificação mais complexa para a adaptação do sinalizador automático, que terá que interpretar a nova glosa com emoções.

Ademais, os resultados do presente estudo são promissores, de modo que levam a crer que esse tipo de adaptação pode ser bem-vista por parte dos usuários, resolvendo algumas das usuais reclamações dos mesmos. Além disso, o estudo também auxilia no preenchimento do *gap* referente a escassez de pesquisas direcionadas para *low-resource languages*, como é o caso da Libras.

Para além, aponta-se como trabalhos futuros a ampliação da base de dados de treinamento do analisador de emoções, bem como a exploração de outras arquiteturas de redes neurais voltadas para o mesmo contexto de uso. Ainda, visa-se o aprimoramento

do serviço de tradução automática adaptado para um melhor tratamento das frases e identificação de múltiplas emoções. Outro ponto que requer maior atenção em trabalhos futuros é a sintaxe de adição de emoções à glosa, que pode ser ampliada com a adição de metadados ou atributos às *tags* para uma melhor apresentação da emoção. Quanto ao sinalizador automático, aponta-se a possibilidade de ampliação das animações de expressões faciais do avatar 3D para um melhor paralelo entre as expressões faciais das emoções básicas humanas. Por fim, é apontada a necessidade de testes com usuários reais para uma validação mais concisa do sistema como um todo.

REFERÊNCIAS

[1] Almasoud, Ameera M e Hend S Al-Khalifa: *A proposed semantic machine translation system for translating Arabic text to Arabic sign language*. Em *Proceedings of the Second Kuwait Conference on e-Services and e-Systems*, página 23. ACM, 2011.

[2] Araújo, Tiago Maritan Ugulino de: *Uma solução para geração automática de trilhas em Língua Brasileira de Sinais em conteúdos multimídia*. 2012.

[3] Bai, Yunhe e David Bruno: *Addressing Communication Barriers Among Deaf Populations Who Use American Sign Language in Hearing-Centric Social Work Settings*. Columbia Social Work Review, 18(1):37–50, 2020.

[4] Bijalwan, Vishwanath, Vinay Kumar, Pinki Kumari e Jordan Pascual: *KNN based machine learning approach for text and document mining*. International Journal of Database Theory and Application, 7(1):61–70, 2014.

[5] Bradbury, James, Stephen Merity, Caiming Xiong e Richard Socher: *Quasi-recurrent neural networks*. arXiv preprint arXiv:1611.01576, 2016.

[6] Brahma, Siddhartha: *Improved Sentence Modeling using Suffix Bidirectional LSTM*. arXiv preprint arXiv:1805.07340, 2018.

[7] Cho, Kyunghyun, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk e Yoshua Bengio: *Learning phrase representations using RNN encoder-decoder for statistical machine translation*. arXiv preprint arXiv:1406.1078, 2014.

[8] Cihan Camgoz, Necati, Simon Hadfield, Oscar Koller, Hermann Ney e Richard Bowden: *Neural sign language translation*. Em *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, páginas 7784–7793, 2018.

[9] Conneau, Alexis, Holger Schwenk, Loïc Barrault e Yann Lecun: *Very deep convolutional networks for text classification*. arXiv preprint arXiv:1606.01781, 2016.

[10] Coppin, Géraldine e David Sander: *Theoretical approaches to emotion and its measurement*. Em *Emotion measurement*, páginas 3–30. Elsevier, 2016.

[11] Corrêa, Ygor, Rafael Peduzzi Gomes e Carina Rebello Cruz: *A desambiguação de palavras homônimas em sentenças por aplicativos de Tradução Automática Português Brasileiro-Libras*. Trabalhos em Linguística Aplicada, 57(1):319–351, 2018.

[12] Devlin, Jacob, Ming Wei Chang, Kenton Lee e Kristina Toutanova: *Bert: Pre-training of deep bidirectional transformers for language understanding*. arXiv preprint arXiv:1810.04805, 2018.

[13] Dizeu, Liliane Correia Toscano de Brito e Sueli Aparecida Caporali: *A língua de sinais constituindo o surdo como sujeito.* Educação & Sociedade, 26(91):583–597, 2005.

[14] Dosciatti, Mariza Miola, LPC Ferreira e EC Paraiso: *Identificando emoçoes em textos em português do brasil usando máquina de vetores de suporte em soluçao multiclasse.* ENIAC-Encontro Nacional de Inteligência Artificial e Computacional. Fortaleza, Brasil, 2013.

[15] Eisenschlos, Julian, Sebastian Ruder, Piotr Czapla, Marcin Kardas, Sylvain Gugger e Jeremy Howard: *MultiFiT: Efficient Multi-lingual Language Model Fine-tuning.* arXiv preprint arXiv:1909.04761, 2019.

[16] Ekman, Paul e Wallace V Friesen: *Manual for the facial action coding system.* Consulting Psychologists Press, 1978.

[17] Emmorey, Karen: *Language, cognition, and the brain: Insights from sign language research.* Psychology Press, 2001.

[18] Faris, Hossam, Ibrahim Aljarah e Seyedali Mirjalili: *Training feedforward neural networks using multi-verse optimizer for binary classification problems.* Applied Intelligence, 45(2):322–332, 2016.

[19] Gago, Jennifer J, Valentina Vasco, Bartek Łukawski, Ugo Pattacini, Vadim Tikhanoff, Juan G Victores e Carlos Balaguer: *Sequence-to-Sequence Natural Language to Humanoid Robot Sign Language.* arXiv preprint arXiv:1907.04198, 2019.

[20] Goldberg, Yoav: *Neural network methods for natural language processing.* Synthesis Lectures on Human Language Technologies, 10(1):1–309, 2017.

[21] Hanke, Thomas: *HamNoSys-representing sign language data in language resources and language processing contexts.* Em *LREC*, volume 4, páginas 1–6, 2004.

[22] Hanke, THOMAS: *HamNoSys–Hamburg Notation System for Sign Languages.* Institute of German Sign Language, Accessed in, 7, 2010.

[23] Hirschberg, Julia e Christopher D Manning: *Advances in natural language processing.* Science, 349(6245):261–266, 2015.

[24] Hochreiter, Sepp e Jürgen Schmidhuber: *Long short-term memory.* Neural computation, 9(8):1735–1780, 1997.

[25] Howard, Jeremy e Sebastian Ruder: *Universal language model fine-tuning for text classification.* arXiv preprint arXiv:1801.06146, 2018.

[26] Jagdale, Rajkumar S, Vishal S Shirsat e Sachin N Deshmukh: *Sentiment analysis of events from Twitter using open source tool*. IJCSMC, 5(4):475–485, 2016.

[27] Jagtap, VS e Karishma Pawar: *Analysis of different approaches to sentence-level sentiment classification*. International Journal of Scientific Engineering and Technology, 2(3):164–170, 2013.

[28] Johnson, Rie e Tong Zhang: *Deep pyramid convolutional neural networks for text categorization*. Em *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, páginas 562–570, 2017.

[29] Khurana, Diksha, Aditya Koli, Kiran Khatter e Sukhdev Singh: *Natural language processing: State of the art, current trends and challenges*. arXiv preprint arXiv:1708.05148, 2017.

[30] Ko, Sang Ki, Chang Jo Kim, Hyedong Jung e Choongsang Cho: *Neural sign language translation based on human keypoint estimation*. Applied Sciences, 9(13):2683, 2019.

[31] Lemos, Felipe Herminio *et al.*: *Uma proposta de protocolo de codificação de libras para sistemas de tv digital*. 2012.

[32] Li, Jiwei, Xinlei Chen, Eduard Hovy e Dan Jurafsky: *Visualizing and understanding neural models in nlp*. arXiv preprint arXiv:1506.01066, 2015.

[33] Loper, Edward e Steven Bird: *NLTK: the natural language toolkit*. arXiv preprint cs/0205028, 2002.

[34] Lopez, Marc Moreno e Jugal Kalita: *Deep Learning applied to NLP*. arXiv preprint arXiv:1703.03091, 2017.

[35] McCleary, Leland e Evani Viotti: *Transcrição de dados de uma língua sinalizada: um estudo piloto da transcrição de narrativas na língua de sinais brasileira (LSB)*. Bilinguismo e surdez. Questões linguísticas e educacionais. Goiânia: Cânone Editorial, páginas 73–96, 2007.

[36] Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S Corrado e Jeff Dean: *Distributed representations of words and phrases and their compositionality*. Em *Advances in neural information processing systems*, páginas 3111–3119, 2013.

[37] Miller, George A: *WordNet: a lexical database for English*. Communications of the ACM, 38(11):39–41, 1995.

[38] Moraes, Rodrigo, JoãO Francisco Valiati e Wilson P GaviãO Neto: *Document-level sentiment classification: An empirical comparison between SVM and ANN*. Expert Systems with Applications, 40(2):621–633, 2013.

[39] Mosolova, Anna, Vadim Fomin e Ivan Bondarenko: *Text Augmentation for Neural Networks.* Em *AIST (Supplement)*, páginas 104–109, 2018.

[40] Niedenthal, Paula M e François Ric: *Psychology of emotion.* Psychology Press, 2017.

[41] Oliveira, Caio César Moraes de, Thaís Gaudencio do Rêgo, Manuella Aschoff Cavalcanti Brandão Lima e Tiago Maritan Ugulino de Araújo: *Analysis of rule-based machine translation and neural machine translation approaches for translating portuguese to LIBRAS.* Em *Proceedings of the 25th Brazillian Symposium on Multimedia and the Web*, páginas 117–124, 2019.

[42] Oliveira, Tiago, Paula Escudeiro, Nuno Escudeiro, Emanuel Rocha e Fernando Maciel Barbosa: *Automatic sign language translation to improve communication.* Em *2019 IEEE Global Engineering Education Conference (EDUCON)*, páginas 937–942. IEEE, 2019.

[43] Omara, Eslam, Mervat Mosa e Nabil Ismail: *Emotion Analysis in Arabic Language Applying Transfer Learning.* Em *2019 15th International Computer Engineering Conference (ICENCO)*, páginas 204–209. IEEE, 2019.

[44] Paiva, Valeria de, Alexandre Rademaker e Gerard de Melo: *OpenWordNet-PT: An Open Brazilian Wordnet for Reasoning.* Em *Proceedings of COLING 2012: Demonstration Papers*, páginas 353–360, Mumbai, India, dezembro 2012. The COLING 2012 Organizing Committee. `http://www.aclweb.org/anthology/C12-3044`, Published also as Techreport http://hdl.handle.net/10438/10274.

[45] Peixoto, Renata Castelo: *Entre palavras e sinais: algumas considerações sobre a alfabetização em língua portuguesa de alunos surdos.* 2019.

[46] Pennington, Jeffrey, Richard Socher e Christopher D Manning: *Glove: Global vectors for word representation.* Em *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, páginas 1532–1543, 2014.

[47] Plutchik, Robert: *The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice.* American scientist, 89(4):344–350, 2001.

[48] Quadros, R. M. e L. B. Karnopp: *Língua de sinais brasileira: estudos linguísticos.* Artmed, Porto Alegre:RS, 2004.

[49] Quadros, Ronice Muller de, Aline Lemos Pizzio e Patrícia Luiza Ferreira Rezende: *Língua Brasileira de Sinais I.*

[50] Rocha, Cleomar e Sarah Caetano de Melgaço: *O uso de aplicativos para tradução de Libras*. Em *Anais do V Simpósio Internacional de Inovação em Mídias Interativas*, 2018.

[51] Rocha Costa, Antônio Carlos da e Graçaliz Pereira Dimuro: *SignWriting-Based Sign Language Processing*. Em *International Gesture Workshop*, páginas 202–205. Springer, 2001.

[52] Rodrigues, Hugo e Fábio Libório Rocha: *Uma definição constitutiva de emoções: a constitutive definition of emotions*. Revista Húmus, 5(15), 2016.

[53] Rosa Zucolotto, Marcele Pereira da, Luciana Rodrigues Ruiz e Najara Ferrari Pinheiro: *REFLEXÕES SOBRE LINGUAGEM, SOCIEDADE E SURDEZ*. Revista Uniabeu, 12(30):134–147, 2019.

[54] Ryan, Claire e Paige Johnson: *Understanding Language Deprivation and Its Role in Deaf Mental Health*. American Annals of the Deaf, 164(4):519–524, 2019.

[55] Sailunaz, Kashfia, Manmeet Dhaliwal, Jon Rokne e Reda Alhajj: *Emotion detection from text and speech: a survey*. Social Network Analysis and Mining, 8(1):28, 2018.

[56] San-Segundo, Rubén, Juan Manuel Montero, R Córdoba, Valentin Sama, F Fernández, LF D'Haro, Verónica López-Ludeña, D Sánchez e Antonio García: *Design, development and field evaluation of a Spanish into sign language translation system*. Pattern Analysis and Applications, 15(2):203–224, 2012.

[57] Sander, David: *Models of emotion*. The Cambridge handbook of human affective neuroscience, páginas 5–56, 2013.

[58] Saritas, Mucahid Mustafa e Ali Yasar: *Performance Analysis of ANN and Naive Bayes Classification Algorithm for Data Classification*. International Journal of Intelligent Systems and Applications in Engineering, 7(2):88–91, 2019.

[59] Silva, Emely Pujólli da, Paula Dornhofer Paro Costa, Kate Mamhy Oliveira Kumada, José Mario De Martino e Gabriela Araújo Florentino: *Recognition of affective and grammatical facial expressions: a study for Brazilian sign language*. Em *European Conference on Computer Vision*, páginas 218–236. Springer, 2020.

[60] Singh, Vivek Kumar, Rajesh Piryani, Ashraf Uddin e Pranav Waila: *Sentiment analysis of movie reviews: A new feature-based heuristic for aspect-level sentiment classification*. Em *2013 International Mutli-Conference on Automation, Computing, Communication, Control and Compressed Sensing (iMac4s)*, páginas 712–717. IEEE, 2013.

[61] Souza, Dyliane Mourí Silva de: *O Efeito do Sentimento do Investidor Manifesto via Twitter sobre o Comportamento dos Retornos e do Volume Negociado no Mercado Acionário Brasileiro.* Tese de Mestrado, Universidade Federal da Paraíba, 2020.

[62] Stoll, Stephanie, Necati Cihan Camgöz, Simon Hadfield e Richard Bowden: *Sign Language Production using Neural Machine Translation and Generative Adversarial Networks.* Em *BMVC*, página 304, 2018.

[63] Sun, Chi, Xipeng Qiu, Yige Xu e Xuanjing Huang: *How to fine-tune BERT for text classification?* Em *China National Conference on Chinese Computational Linguistics*, páginas 194–206. Springer, 2019.

[64] Sutskever, Ilya, Oriol Vinyals e Quoc V Le: *Sequence to sequence learning with neural networks.* Em *Advances in neural information processing systems*, páginas 3104–3112, 2014.

[65] Tabassum, Huma e Sohaib Ahmed: *EmotiOn: An Ontology for Emotion Analysis.* Em *1st National Conference on Emerging Trends and Innovations in Computing and Technology, Karachi, Pakistan*, 2016.

[66] Valli, Clayton e Ceil Lucas: *Linguistics of American sign language: an introduction.* Gallaudet University Press, 2000.

[67] Wairagya, Ida Bagus Nyoman, Putu Wira Buana e I Made Sukarsa: *Development of English-to-Sign-Language Translation System on Android.* International Journal of Computer Engineering and Information Technology, 11(7):157–163, 2019.

[68] Wu, Felix, Angela Fan, Alexei Baevski, Yann N Dauphin e Michael Auli: *Pay Less Attention with Lightweight and Dynamic Convolutions.* arXiv preprint arXiv:1901.10430, 2019.

[69] Yegnanarayana, Bayya: *Artificial neural networks.* PHI Learning Pvt. Ltd., 2009.

ANEXO A - Artigo Publicado no WebMedia 2020

Using Data Augmentation and Neural Networks to Improve the Emotion Analysis of Brazilian Portuguese Texts

Vinícius Veríssimo
LAVID/UFPB
vinicius.matheus@lavid.ufpb.br

Rostand Costa
LAVID/UFPB
rostand@lavid.ufpb.br

ABSTRACT

Information and Communication Technologies present as an interesting alternative for the mitigation of barriers that arise in the context of communication of information, mainly as technologies aimed at the machine translation of content in oral language into sign language. After years, despite the improvement of these technologies, the use of them still divides the opinions of the Deaf Community, due to the low emotional expressiveness of 3D avatars. Therefore, as a way to assist the machine translation of texts in oral language to sign language, this study aims to evaluate the influence of the parameters of a data augmentation method in a textual dataset and the use of neural networks for emotion analysis of Bazilian Portuguese texts. The analysis of emotions in texts presents a relevant challenge in diversity due to the nuances and different forms of expression that the human language uses. In this context, the use of deep neural networks has gained enough space as a way to deal with these challenges, mainly with the use of algorithms that deal with emotion analysis as a textual classification task, such as the MultiFiT approach. To circumvent the scarcity of data in Brazilian Portuguese aimed at this task, some strategies for increasing data were evaluated and applied to improve the database used in training. The results of the emotion analysis experiments with Transfer Learning pointed to accuracy above 94% in the best case.

CCS CONCEPTS

• **Human-centered computing** → **Accessibility technologies**; *Accessibility systems and tools*; • **Computing methodologies** → **Neural networks**.

KEYWORDS

emotion analysis, neural networks, accessibility, sign language

ACM Reference Format:
Vinícius Veríssimo and Rostand Costa. 2020. Using Data Augmentation and Neural Networks to Improve the Emotion Analysis of Brazilian Portuguese Texts. In *Brazilian Symposium on Multimedia and the Web (WebMedia '20), November 30-December 4, 2020, São Luís, Brazil.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3428658.3431080

WebMedia '20, November 30-December 4, 2020, São Luís, Brazil

ACM ISBN 978-1-4503-8196-3/20/11…$15.00
https://doi.org/10.1145/3428658.3431080

1 INTRODUÇÃO

As trocas de informações que ocorrem no âmbito mundial são, majoritariamente, voltadas para os usuários de línguas orais (LO), ou seja, a comunicação é feita, em grande parte, por meio de áudios, textos e vídeos voltados para pessoas ouvintes. Esse fenômeno também ocorre no Brasil, fazendo assim com que uma parcela considerável da população, composta pela população de pessoas surdas, fique à margem do sistema de comunicação.

No Brasil, segundo o censo de 2010 do IBGE [1], por volta de 10 milhões de pessoas possuem algum grau de perda auditiva grave (cerca de 5% da população). Dentre essas pessoas, 2,7 milhões possuem um grau de surdez absoluta. Ou seja, é uma relevante parcela da população que está suscetível a sofrer os efeitos de viverem em um meio onde a maioria das comunicações e distribuição de informações ocorre por meio de línguas orais.

O contato exclusivo com línguas orais é de extrema dificuldade para a população surda, visto que os surdos têm como língua natural a Língua de Sinais (LS). A privação de conteúdo em LS é constantemente relacionado com resultados negativos nos processos de desenvolvimento cognitivo e de aprendizado da surdos [2, 32], o que impacta diretamente em como uma pessoa surda se integra e participa de sua comunidade [8], uma vez que a sua capacidade de absorver e difundir informações fica limitada.

Diante de tais desafios, também emergiu uma série de maneiras de tentar minimizar os problemas de comunicação que pessoas surdas enfrentam como, por exemplo, o uso de intérpretes em eventos presenciais e transmissões televisivas e também a disponibilização obrigatória de intérpretes para alunos surdos em escolas e universidades. Entretanto, há contextos em que a dinâmica e a escala tornam praticamente impossível que sejam resolvidos apenas por intérpretes humanos.

Neste sentido, várias frentes de pesquisa buscam soluções capazes de aumentar a capacidade de comunicação entre pessoas surdas e pessoas ouvintes, tais como a utilização de sistemas que possibilitam a presença virtual de intérpretes de língua de sinais numa interação onde uma das partes é uma pessoa surda [2], e a tradução de forma automática [1, 6, 18, 27, 36]. Dentre essas, a tradução automática se sobressai, visto que nem sempre a presença de um intérprete humano é possível ou viável, como por exemplo, na internet, que possui um conteúdo extramente dinâmico e primariamente voltado para usuários ouvintes.

Entretanto, mesmo com o reconhecimento da sua utilidade e adoção crescente, os tradutores automáticos para línguas de sinais enfrentam críticas em alguns aspectos. Rocha e Melgaço [30], realizaram uma pequisa de opinião com usuários de aplicativos de tradução automática de Português Brasileiro para LIBRAS (Língua

[1] https://censo2010.ibge.gov.br/

Brasileira de Sinais). Constatou-se que há uma visão positiva por parte da comunidade surda sobre o uso dessas tecnologias, todavia, a expressividade do avatar 3D é citado como um ponto a ser melhorado. Especificamente, aspectos como a interpretação do avatar e a falta de humanidade do mesmo, bem como a necessidade de expressão de sentimentos e expressões faciais e corporais adequadas foram ditos como as principais barreiras. Entretanto, a pesquisa também apontou que os entrevistados acreditam que tais problemas podem ser sanados para elevar a aceitação dessas ferramentas dentro da comunidade surda.

Diante do exposto, este trabalho tem como objetivo avaliar a influência dos parâmetros de uma estratégia de *data augmentation* em uma base de dados textual e do uso de redes neurais para a análise de emoções em textos em Português Brasileiro. A proposta é avaliada na utilização com redes neurais profundas através do método MutiFiT [12], o qual aplica o conceito de *Transfer Learning* para o treinamento das redes neurais. Dessa maneira, visa-se a sua utilização, em estudos futuros, como forma de subsidiar a melhoria da expressividade dos avatares 3D usados em tradutores automáticos Português-LIBRAS.

Os resultados encontrados demostram uma influência positiva na utilização de *data augmentation* e uma estratégia baseada em rede neurais para a tarefa de análise de emoções em textos, uma vez que, fora encontrada uma acurácia média de 94,29%, no seu melhor caso, consideravelmente superior ao à acurácia de 61% obtida por Dosciatti, Ferreira e Paraiso [11] com a mesma base de dados.

O restante do documento está organizado como segue. Na Seção 2 são apresentados conceitos da análise de emoções e alguns dos métodos mais conhecidos para se trabalhar com análise de emoções em textos. Na Seção 3 são apresentados alguns trabalhos relacionados com a pesquisa. Na Seção 4 será apresentada uma proposta de utilização do analisador de emoções em conjunto com um tradutor automático. Na Seção 5 é apresentada a metodologia utilizada para a realização dos experimentos de *data augmentation* da base de dados selecionada para a pesquisa. Na Seção 6 é apresentada uma análise dos resultados dos experimentos e, finalmente, na Seção 7 são feitas as considerações finais do trabalho.

2 ANÁLISE DE EMOÇÕES

Segundo Rodrigues e Rocha [31] e Coppin e Sander [7], não há ainda um consenso na literatura sobre o que são as emoções, cada área de estudo tem seus próprios conjuntos de conceitos. Por exemplo, Coppin e Sander [7] são favoráveis a definição de Sander [34] de que a emoção se baseia em eventos, que consiste em "um mecanismo de elicitação emocional baseado em relevância que molda uma resposta multi-emocional". Nesse sentido, a resposta pode ser, por exemplo, uma tendência de ação, uma reação automática, expressão ou sentimento.

De forma semelhante, Niedenthal e Ric [26] definem emoções como o fogo que alimenta o comportamento humano e as forças motivadoras da vida. Logo, exemplificando, o processo de sentir uma emoção como, por exemplo, o *medo*, requer a percepção de algo no mundo, lembrar que aquilo é uma ameaça e realizar uma ação de fuga.

Já como uma forma de identificar as emoções humanas, Plutchik [29] criou um modelo que identifica 8 emoções primárias, sendo essas: raiva, antecipação, alegria, confiança, medo, surpresa, tristeza e nojo. A Figura 1 apresenta o modelo que é graficamente definido por um conjunto de círculos circunvizinhos, de modo que o círculo do meio é composto pelas emoções primárias e, os círculos interno e externo, pela menor e maior intensidade dessas emoções, respectivamente. As demais emoções seriam misturas das emoções primárias.

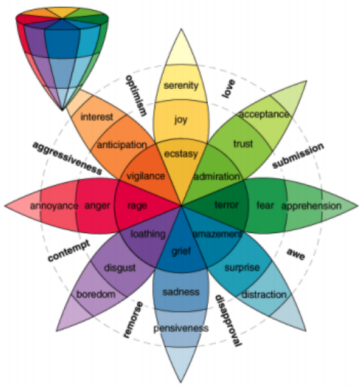

Figura 1: Círculo de emoções proposto por Plutchik [29]

Um exemplo para compreender o modelo de Plutchik é observar a emoção de alegria (*joy*), que em maior intensidade se manifesta como êxtase (*ecstasy*) e em menor intensidade como serenidade (*serenity*), e quando misturada com confiança (*trust*), se manifesta como amor (*love*). A representação se assemelha à da roda de cores, de tal forma que as emoções opostas estão em um ângulo de 180º como cores complementares. Além disso, as intensidade das emoções são associadas com a intensidades das cores.

Assim como no modelo de Plutchik [29], a área de *detecção/análise de emoções*, referenciada também como *detecção/análise de sentimentos* ou *mineração de opinião* (do inglês *opinion mining*), é uma forma mais robusta que busca estudar a identificação/detecção das emoções humanas, além de as investigar. Alguns autores diferenciam a análise de sentimentos da análise de emoções [16, 28, 33], muito embora elas sejam citadas diversas vezes como sinônimos.

A análise de emoções é significativamente uma tarefa mais complexa que a análise do sentimento [33], isso porque a análise de sentimentos é tida como sendo apenas o estudo da polaridade do sentimento (positivo ou negativo), já a análise de emoções abrange diferentes classes de emoções (e.g. alegria, tristeza, medo, etc...), podendo até envolver a identificação da intensidade dessas emoções [16, 28, 33].

Tanto a análise de emoções quanto a análise de sentimentos podem ser realizadas através de diferentes tipos de dados, tais como imagem, vídeo, áudio e texto, e podem ser utilizados em procedimentos automatizados ou semi-automatizados [33]. Dentre as fontes de dados potenciais, pode-se imaginar que a análise de textos parece ser a mais simples em um primeiro momento, entretanto, a

grande dimensionalidade de representação das línguas humanas torna tal tarefa extremamente complexa.

Além de informações, textos também são capazes de expressar a opinião e emoção do autor [11], mas na maior parte do tempo as emoções são expressas de forma subentendida ou compacta, podendo ser interpretada de forma ambígua, dadas as características das línguas humanas [33].

Em face da complexidade de sentimentos e emoções que podem estar representado em um texto, esse tipo de análise pode ocorrer em três níveis: *nível de documento*, *nível de sentença* e *nível de aspecto*. Sendo que o último nível também é conhecido como *nível de característica* (do inglês *Feature Level*), no qual se busca a análise de um termo específico e não do documento, sentença ou frase [16, 35].

A automação da análise de emoções em textos se destaca por poder ser explorada de diversas formas, não apenas como parte do aprimoramento da interação homem-máquina. Áreas como mídias-sociais, negócios, medicina, entre outras têm muito a se beneficiar desse ramo de pesquisa. Por isso, existem diversas estratégias para diferentes contextos de uso da análise de sentimentos e emoções.

Apesar da complexidade da tarefa, técnicas de aprendizagem de máquina podem ser utilizadas na construção de analisadores [11]. Algumas dessas técnicas são: Naïve Bayes, um algoritmo de aprendizagem probabilística que se baseia no Teorema de Bayes; *Support Vector Machine* (SVM), que tem uma forte base na Teoria do Aprendizado Estatístico; *K-Nearest Neighbors*, que realiza a categorização no plano vetorial, levando em consideração a categoria do k vizinhos mais próximos [3]; e mais recentemente as Redes Neurais.

As redes neurais (NNs - do inglês, *Neural Networks*) são um tipo de modelo computacional para a realização de tarefas de reconhecimento de padrões e inspiradas na estrutura e no funcionamento das nossas redes neurais biológicas [38]. Normalmente são representadas como um conjunto de nós conectados, arranjados em forma de rede e agrupados em camadas, geralmente divididas em camadas de entrada, ocultas e de saída. Cada conexão de nós está associada a um peso, que é calculado durante a fase de treinamento [24]. Dependendo da dimensionalidade dessas camadas, essas redes recebem a nomenclatura de Redes Neurais Profundas (DNN - do inglês, *Deep Neural Networks*)

Dentre as técnicas de aprendizagem de máquina supracitadas, as Redes Neurais se destacam, uma vez que elas atingiram o estado da arte em diversos contextos de uso, e na análise de emoções e sentimentos não é diferente. Na próxima seção será explorado como a análise de emoções e sentimentos de textos é tratada pelas Redes Neurais. Em adição, é apresentado o método MultiFiT [12], que serve como base para a presente pesquisa.

2.1 Análise de Emoções em Texto com Redes Neurais

O uso de Redes Neurais representa uma poderosa e atrativa ferramenta para o processamento de linguagem natural (PLN) [14], visto que muitos modelos de redes neurais atingem ou até mesmo superam o estado da arte em tarefas do PLN quando comparados com outras estratégias de aprendizagem de máquina [19].

Exitem dois tipos de redes neurais que são majoritariamente utilizadas no processamento de linguagens naturais, são elas: Redes *Feed-forward* e Redes Recorrentes (RNN) (do inglês, *Recurrent Neural Networks*) [14].

As Redes *Feed-forward* são o tipo mais comum de NN. Elas se caracterizam por organizarem os neurônios em camadas, essencialmente os organizando em camada de entrada, camadas ocultas e camada de saída [13].

Já as Redes Recorrentes (RNN) foram feitas para trabalharem com informações sequenciais. Elas são chamadas de recorrentes por aplicarem o mesmo procedimento em todos os elementos da sequência de entrada. As RNNs são conhecidas pelo seu mecanismo de memória, onde os elementos anteriores já computados da sequência também influenciam na computação dos próximos elementos [22].

De forma geral, a sequência recebida pela RNN pode conter um tamanho variável e a rede produzirá um vetor de tamanho fixo que sumariza toda a informação da sequência [14]. Um texto pode ser intuitivamente visualizado como um sequência de *tokens* (letras ou palavras) com dependência temporal, ou seja, um *tokens* na posição t pode influenciar o sentido de um outro *tokens* em $t + n$, assim como pode ser influenciado por um *tokens* em $t - n$.

Tão importante quanto os tipos de redes neurais utilizadas na construção dos sistemas é a forma como essas redes estão arquiteturalmente, em tamanho, quantidade e densidade das camadas, e qual método de treinamento utilizado. Uma parte considerável dos trabalhos sobre DNNs exploram exatamente isso, buscando alternativas de redes neurais e métodos de treinamento que sirvam a um propósito específico ou geral.

Muitos dos problemas da área de PLN fazem parte de um conjunto maior de problemas que são o foco dessas soluções. Nesse sentido, a análise de sentimentos/emoções pode ser tratada como um problema de classificação de textos, onde as classes dos textos são as próprias emoções. Nesse caso, alguns métodos destinados para essa classe de problema também podem ser utilizadas para esse propósito.

Um exemplo de utilização de NN de classificação de textos é com o método MultiFiT (*Multi-lingual Fine-Tuning*), de Eisenschlos et al. [12], que foi criado para otimizar a ideia trazida pela ULMFiT [15] para a utilização de classificação de textos em qualquer idioma. Ambas arquiteturas atingiram excelentes resultados em seus testes, e apesar do MultiFiT não se manter como estado da arte, ela é uma solução mais leve que as demais, e possui mais recursos de fácil acesso para diversas línguas, incluindo o Português.

O MultiFiT utiliza a estratégia de *transfer learning* na sua construção. A ideia geral é utilizar mais de uma rede neural de modo que se reduza a complexidade do problema a cada rede utilizada. No caso do MultiFiT são utilizadas rede de *Language Model* (LM) e classificação.

A Figura 2 apresenta uma visão do processo de treinamento do MultiFiT. De forma geral, o primeiro passo (1) é o treinamento de uma rede LM bidirecional com uma base de dados de propósito geral da língua alvo, nessa rede LM será realizado o procedimento de *fine-tuning* (ou afunilamento); (2) é então feita uma especialização da rede neural com a base de dados que se está trabalhando, porém sem as classes, apenas os textos de entrada; a próxima etapa (3) é utilizar o conhecimento adquirido pela rede LM, ou seja, os pesos da rede,

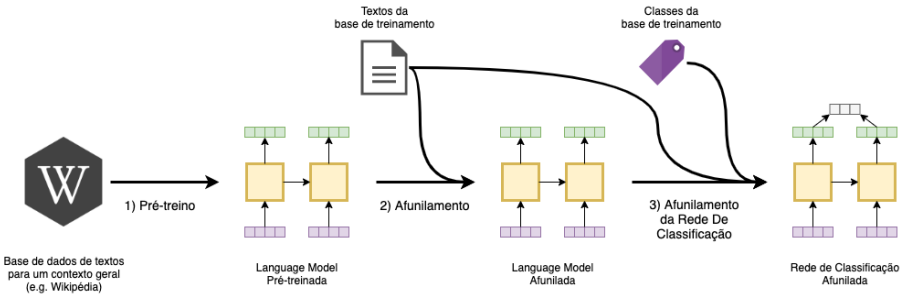

Figura 2: Etapas da arquitetura MultiFiT

e utilizá-lo no processo de afunilamento da rede de classificação, usando a mesma base de dados, mas agora com as classes dos textos.

Na definição original da ULMFiT são utilizadas redes de tipo AWD-LSTM (*ASGD Weight-Dropped Long Short-Term memory*) em todo o processo, já o MultiFiT se utiliza de redes QRNN (*Quasi-Recurrent Neural Networks*) [4], um tipo de rede convolucional que pode ser paralelizada e que utiliza funções de *pooling* recorrentes. Essas redes superam a performance das redes LSTM tradicionais e são computacionalmente mais rápidas e leves.

Existem muitas formas de se utilizar DNN na análise de sentimentos e emoções em texto [5, 10, 17, 37]. Alguns desses trabalhos foram ou atingiram o estado da arte para essa tarefa e mostram a eficiência na utilização de redes neurais para a análise de sentimentos e emoções, mesmo que nem sempre as soluções sejam diretamente projetadas para isso, mas a redução desse problema para uma tarefa de classificação de textos possibilita esses resultados.

Na próxima seção serão apresentados alguns trabalhos relacionados com o tema e objetivo deste estudo.

3 TRABALHOS RELACIONADOS

Howard e Ruder [15] propõem um método de transferência de aprendizado que pode ser aplicada para qualquer tipo de tarefa de processamento de linguagem natural. O método ULMFiT (*Universal Language Model Finetuning*) consiste em utilizar uma rede neural para *Language Model* que é capaz de compreender as dependências da palavras da língua que está sendo trabalhada, diminuindo a complexidade necessária para a construção da rede responsável pelo processo de PLN mais específico.

O processo proposto por Howard e Ruder [15] consiste em utilizar uma rede de LM pré-treinada para um contexto geral, especializá-la para o contexto que está sendo trabalhado, e utilizar o conhecimento adquirido em uma rede especializada para a tarefa de PLN desejada através de *tranfer learning*. Como experimento, o método proposto foi utilizado em tarefas de classificação de textos (análise de sentimentos, classificação de questões e classificação de tópicos). A proposta conseguiu atingir o estado-da-arte para as base de dados utilizadas. Esse método serviu como base para o método MultiFiT.

Eisenschlos [12] se baseia no ULMFiT para a definição de um método de *transfer learning* para tarefas de PLN, o MultiFit. Apesar do ULMFiT ter sido pensada para ser utilizada com qualquer língua, o MultiFit aperfeiçoa isso levando em consideração que nem todas as línguas possuem construção de palavras semelhantes ao Inglês, sendo assim, enquanto que o ULMFiT limita os dados de entrada à granularidade de palavras, o MultiFit possibilita a utilização de sub-palavras. Outra diferença está na utilização de redes QRNN em todo o processo, em vez das AWD-LSTM utilizadas no ULMFiT, o que possibilita mais paralelização do treinamentos das redes, além das QRNN serem mais leves e rápidas que as redes AWD-LSTM. No trabalho, o MultiFiT supera o ULMFiT nas tarefas de classificação, principalmente quando comparados na utilização com diferentes tipos de línguas. Assim, esse método serve como base para o processo de classificação utilizado nessa pesquisa.

Omara, Mosa e Ismail [28] estudaram a utilização de redes neurais convolucionais treinadas utilizando *transfer learning* a fim de realizar detecção de emoções em textos em árabe. Sobre a base de dados utilizada foi aplicada uma técnica de *data augmentation* através da substituição de sinônimos com o propósito de deixar a rede neural mais robusta. A estratégia proposta atingiu o estado da arte para a classificação de emoções em textos em árabe. Esse trabalho se relaciona à pesquisa em demonstrar a utilização de *data augmentation* no treinamento de uma rede neural que se utiliza de *transfer learning*.

O artigo de Dosciatti, Ferreira e Paraiso [11] utiliza Máquina de Vetores de Suporte (SVM) na identificação das seis emoções básicas (alegria, tristeza, raiva, medo, desgosto e surpresa) em textos escritos em Português do Brasil. Para testar o método proposto, um corpus formado por 1.750 textos foi construído. Destaca-se ainda que o método não utiliza nenhum recurso linguístico adicional, como um léxico especialmente preparado para relacionar palavras ligadas à emoções. A base de dados utilizada nesse trabalho é a mesma utilizada nessa pesquisa.

Na próxima seção será apresentada a proposta de utilização da análise de emoções em conjunto com a tradução automática para língua de sinais.

4 UM ANALISADOR DE EMOÇÕES PARA USO EM TRADUTORES AUTOMÁTICOS

O processo de análise automática da emoção do texto pode ser realizada de forma paralela à tradução automática. Como mostra a Figura 3, o mesmo texto de entrada do tradutor automático será utilizado pelo classificador. A emoção identificada pelo algoritmo será

então aplicada ao texto traduzido, normalmente em uma notação escrita das línguas de sinais, como a glosa.

O sinalizador automático deverá interpretar o texto traduzido com a informação da emoção e realizar da transmissão de forma adequada, seja ela de forma gráfica, escrita, como expressão facial ou corporal.

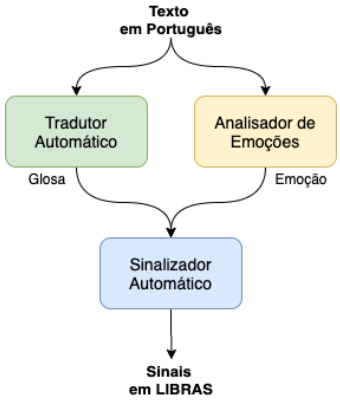

Figura 3: Processo da Análise de Emoções em Conjunto com a Tradução Automática

Um dos principais requisitos para aplicação da estratégia proposta é a disponibilidade de uma base de dados suficientemente grande e abrangente para o treinamento adequado de uma rede neural para a análise de emoções em Português Brasileiro. No melhor do nosso conhecimento, não há ainda uma base em classificada com tais atributos voltada para a análise de emoções.

Para sanar esta lacuna e tendo como ponto de partida uma base de pequena escala que foi encontrada em um busca na literatura, o foco principal deste trabalho é investigar estratégias de torná-la apta a ser utilizada para treinar um modelo MultiFiT.

4.1 Definição da Base de Dados

A base de dados a ser utilizada como ponto de partida é a definida no trabalho de Dosciatti, Ferreira e Paraíso [11] para a utilização na construção de um analisador de emoções utilizando uma estratégia do tipo SVM.

O conjunto de dados próprio é composto de notícias do site *www.globo.com*, de diversas categorias, contando com um total de 1.750 textos, sendo 250 para cada uma das 6 emoções abordadas (alegria, desgosto, medo, raiva, surpresa e tristeza), escolhidas com base no modelo de Plutchik, e 250 para a classe "neutro". O processo de rotulação dos textos foi realizado por duas pessoas: uma especialista em linguística e outra em linguística computacional. A base de dados desse trabalho pode ser adquirida através de uma solicitação no site do projeto [2].

A quantidade de texto rotulados pelos autores [11] é pequena para a utilização com redes neurais. Diante disso, será realizado um processo de *data augmentation* dessa base de dados.

[2]http://www.ppgia.pucpr.br/ paraiso/mineracaodeemocoes/recursos/g1$_o$2.*php*

4.2 Estratégia de Data Augmentation

Um requisito comum às estratégias que se utilizam de NNs é a necessidade de uma base de dados abrangente o suficiente para a solução do problema. Existem pouquíssimas bases de dados em Português Brasileiro voltadas para a classificação de emoções. Ao contrário de bases de dados para análise de polaridade de sentimento.

Em casos onde a quantidade de dados disponíveis não é suficiente para o treinamento de um algoritmo de aprendizagem de máquina, é comum a aplicação de técnicas de *data augmentation* para a ampliação de exemplos da base de dados.

A realização de *data augmentation* é muito comum em treinamento de redes neurais que utilizam imagem como dados de treinamento. Em suma, essa técnica consistem em um processo que modifica um dado existente na base de dados de treinamento para que ele se torne um novo dado também relevante para a base de treinamento. No contexto de imagens, são comuns processos como rotação, espelhamento, mudança dos canais de cores etc. Já no contexto de dados textuais, modificações semelhantes não podem ser realizadas, pois mudariam o dado a ponto de os deixar errados e inválidos para o aprendizado do algoritmo.

A heurística mais comum para realizar *data augmentation* de textos é substituição de sinônimos. Por exemplo, a heurística de Mosolova, Fomin e Bondarenko [25], consiste em escolher as palavras no texto que podem possuir sinônimos, a partir da identificação das classes gramaticais dessas palavras, como por exemplo por meio de verbos, substantivos e adjetivos, já as palavras das demais classes gramaticais permanecem intactas. Com as palavras de interesse identificadas é necessário listar os sinônimos delas, para isso se utiliza a Wordnet[23], que é um conjunto de dados que contém informações e relacionamento das palavras da língua inglesa, para qual foi originalmente desenvolvida, mas que depois foi implementada em outras línguas, incluindo o Português Brasileiro [9]. A escolha de quais dessas palavras serão substituídas e quais sinônimos serão escolhidos é feito de forma parametrizada e randômica.

Na próxima seção será apresentado o projeto dos experimentos para a definição da melhor estratégia realização do *data augmentation* na base de dados de partida.

5 PROJETO DE EXPERIMENTOS (DOE)

Os experimentos planejados têm como objetivo principal encontrar a melhor combinação de valores dos fatores definidos para realizar o processo de *data augmentation* proposto por Mosolova, Fomin e Bondarenko [25]. A heurística de [25] possui alguns parâmetros que podem ser explorados nos experimentos. Adicionalmente é proposto um novo parâmetro para controlar o balanceamento da base de dados após o *data augmentation*.

Para a definição dos experimentos será utilizado o Planejamento de Experimentos Fatorial 2^k, onde são definidos k fatores de testes que podem influenciar o resultado final. Para cada fator, são definidos dois valores (ou níveis). A partir disso, são planejados 2^k experimentos distintos variando-se os níveis dos fatores.

Para a análise dos resultados foi realizada uma Análise da Variância (ANOVA) para estudar a influência e significância dos fatores nas métricas computacionais geradas. Outro passo da análise foi o cálculo do coeficiente de determinação (R^2) para se verificar a relevância do modelo sobre os resultados alcançados.

5.1 Fatores

A heurística de *data augmentation* através da substituição de sinônimos proposta por Mosolova, Fomin e Bondarenko [25] possui dois parâmetros para controlar como deve ser realizado o processo de *augmentation* dos textos, sendo eles: porcentagem de palavras substituídas por iteração no texto (PctWords) e quantidade máxima de novos textos gerados por texto de origem (MaxNew).

O primeiro parâmetro (porcentagem de palavras substituídas por iteração no texto) diz respeito à porcentagem de palavras válidas para o processo de substituição que poderão ser substituídas a cada iteração num mesmo texto. Isso quer dizer que a cada processo de *augmentation* em um texto da base de dados, apenas uma porcentagem definida das palavras candidatas para substituição serão efetivamente substituídas, tendo seus sinônimos escolhidos aleatoriamente. Para os experimentos, apenas verbos, substantivos e adjetivos serão consideradas palavras válidas para substituição.

O segundo parâmetro (quantidade máxima de novos textos gerados por texto de origem) serve como uma condição de parada do procedimento anterior. Uma outra forma do procedimento anterior parar é caso não seja possível a geração de novos textos, ou seja, qualquer que sejam os sinônimos escolhidos não é possível gerar um texto distinto dos demais. Esse segundo parâmetro limita superiormente a quantidade de textos gerados.

Como a heurística original pode gerar uma quantidade diferente de novos textos para cada classe, propõem-se também a adição de um novo parâmetro (BALANC). Esse novo parâmetro consiste em uma condição binária definindo se a nova base de dados gerada deverá ser balanceada ou não, ou seja, se todas as classes devem conter a mesma quantidade de exemplos ou não. Caso positivo, a quantidade de textos de cada classe será limitada pela classe com menos exemplos.

Nesse sentido, como definido pelo Planejamento de Experimentos Fatorial 2^k, dois níveis (valores) devem ser definidos para os fatores. Para o fator PctWords são definidos os valores 25% e 75%, sendo o menor nível originado de [25] e o maior nível foi definido como um valor de controle. Para o fator MaxNew são definidos os valores 6 e 10, também seguindo os mesmo critérios do anterior. Já o fator BALANC possui um valor binário, o balanceamento ou não da base de dados.

A partir da base de dados original, serão separados 10% dos exemplos para a criação de uma base de dados de teste e 90% para *data augmentation*. Os exemplos da base de dados de teste foram selecionados aleatoriamente para cada bateria de experimentos, ou seja, todos o experimentos de $R1$ na Tabela 3 foram atingidos com a mesma base de dados de teste, assim como em $R2$. Ambas as bases de dados, para teste e para *data augmentation*, iniciam-se balanceadas.

Na Tabela 1 são apresentados os fatores definidos para os experimentos, bem como os seus níveis, são eles: porcentagem de palavras substituídas por iteração no texto (PctWords); quantidade máxima de novos textos por texto de origem (MaxNew); e balanceamento da base de dados (BALANC). Já a Tabela 2 apresenta a tabela de experimentos a serem realizados.

Tabela 1: Níveis dos Fatores

Fator	-1	1
PctWords	25%	75%
MaxNew	6	10
BALANC	Não	Sim

Tabela 2: Matriz de Planejamento

Nº do Experimento	PctWords	MaxNew	BALANC
1	-1	-1	-1
2	-1	-1	1
3	-1	1	-1
4	-1	1	1
5	1	-1	-1
6	1	-1	1
7	1	1	-1
8	1	1	1

5.2 Métricas de Interesse

Os resultados serão avaliados tendo como métrica de interesse a acurácia de acerto das emoções e o *F1-score*, que leva em consideração a precisão e *recall* das predições das redes.

5.3 Configuração do Ambiente

A arquitetura MultiFiT foi implementada utilizando o *framework* fastai[3], uma abstração em auto-nível do *framework* de *machine learning* PyTorch[4]. Como indicado no trabalho de Eisenschlos et al. [12], tanto na implementação da rede de *Language Model* quanto de classificação fora utilizadas 4 camadas de redes QRNN com uma dimensionalidade de 1.550 camadas ocultas.

A rede *Language Model* bidirecional utilizada[5] foi pré-treinada utilizando a base de dados de textos em Português do Wikidata[6], que reúne textos de artigos do Wikipédia em diversas línguas, limitando-se o vocabulário a 15.000 *tokens* e utilizando um vetor *embedding* de tamanho 400.

A implementação da heurística de *data augmentation* se deu na linguagem Python, utilizando as bibliotecas SpaCy[7] e NLTK [21], para a identificação das classes gramaticais, e interface para acessar os sinônimos das palavras na Wordnet em Português.

Todos os experimentos foram executados no ambiente do Google Colab[8].

6 RESULTADOS E ANÁLISE

Seguindo os experimentos planejados na Tabela 2, avaliou-se os fatores: porcentagem de palavras substituídas por iteração no texto (PctWords), quantidade máxima de novos textos por texto de origem (MaxNew) e balanceamento da base de dados (BALANC), para a

[3] https://github.com/fastai/fastai/
[4] https://pytorch.org/
[5] Disponível em https://github.com/piegu/language-models
[6] https://www.wikidata.org/
[7] https://spacy.io/
[8] https://colab.research.google.com/

realização de *data augmentation* da base de dados de Dosciatti, Ferreira e Paraíso [11].

Na Tabela 3 é possível observar que o experimento 8 da primeira bateria (PctWord de 75%, MaxWord de 10 e balanceamento ativado), obteve os melhores resultados tanto em acurácia (94,29%) quanto no *F1-score* (94,28%), consideravelmente superior aos resultados obtidos por Dosciatti, Ferreira e Paraíso [11], que foram de 60,7% de acurácia e 60% de *F1-score*, utilizando uma estratégia de SVM.

Tabela 3: Resultados de Acurácia e *F1-score* dos Experimentos

Nº do Experimento	R1		R2	
	Acurácia	F1-score	Acurácia	F1-score
1	68,00%	67,77%	63,43%	63,13%
2	93,14%	93,15%	90,29%	90,28%
3	67,43%	67,00%	62,86%	62,42%
4	93,14%	93,20%	**91,43%**	**91,48%**
5	66,29%	65,45%	60,00%	58,71%
6	93,71%	93,74%	90,86%	90,94%
7	68,00%	67,65%	60,57%	60,83%
8	**94,29%**	**94,28%**	90,86%	90,85%

A Tabela 4 apresenta uma comparação das acurácias e *F1-score* para cada emoção no Experimento 8 da primeira bateria e no trabalho de Dosciatti, Ferreira e Paraíso [11]. Para as emoções "raiva"e "surpresa"o modelo obteve uma acurácia de 100%, mas em ambos os casos o *F1-score* foi de 98,04%. Assim, é possível visualizar que a proposta apresentada superou os resultados encontrados no trabalho de [11] na classificação de todas as classes e obteve valores de *F1-score* superiores a 90% em todos os casos, ou seja, teve uma alta taxa de acerto nas classificações.

Na segunda bateria de experimentos, o experimento 4 (PctWord de 25%, MaxWord de 10 e balanceamento ativado) obteve o melhor resultado do seu conjunto, obtendo uma acurácia de 91,43% e *F1-score* de 91,48%, menor que o experimento 8 da primeira bateria, mas ainda sim superior aos resultados relatados no trabalho que originou a base de dados.

Ademais, é aparente a diferença dos resultados dos experimentos de números pares em relação aos de números ímpares. Nos experimentos pares o único fator que permanece igual é a aplicação do balanceamento na base de dados, enquanto que os experimentos ímpares não possuem balanceamento, ou seja, isso pode apontar o fator do balanceamento como um grande influente nos resultados. Essa suspeita pode ser comprovada através da aplicação de uma ANOVA sobre os resultados dos experimentos, a qual pode ser observada nas Tabela 6 e Tabela 5.

Tanto na Tabela 6 quanto na Tabela 5 é possível comprovar que o fator BALANC é o que possui maior significância estatística, ou seja, ele exerce uma influência considerável nos resultados. Através da análise, também é possível identificar que, no tocante aos outros fatores e as interações entre eles não possuem significância estatística nos resultados, ou seja, a troca de valores deles não possuem influência significativa nos resultados, até mesmo na interações com o fator BALANC, que possui uma maior significância estatística isoladamente do que nas interações com os demais fatores.

Pela análise das acurácias, o modelo obteve um R^2 de *0,9734* e o R^2 ajustado de *0,9501*. Já analisando o *F1-score*, o modelo obteve um R^2 de *0,9747* e o R^2 ajustado de *0,9525*. Ambos os casos apontam que os modelos se adéquam aos dados apresentados.

Tabela 4: Comparação dos Resultados do Experimento 8 da Primeira Bateria (Exp8 (R1)) de Experimentos com os Resultados de Dosciatti, Ferreira e Paraíso [11] (DFP)

Emoção	Acurácia		F1-score	
	Exp8 (R1)	DFP	Exp8 (R1)	DFP
Alegria	92%	45%	93,88%	46%
Desgosto	88%	39%	91,67%	40%
Medo	92%	81%	93,88%	76%
Neutro	96%	50%	90,57%	51%
Raiva	100%	75%	98,04%	75%
Surpresa	100%	81%	98,04%	78%
Tristeza	92%	54%	93,88%	54%
Média	94,29%	61%	94,28%	60%

Tabela 5: ANOVA da Métrica de Acurácia dos Experimentos

Origem	SS Parcial	*df*	*MS*	F	Prob >F
Modelo	0,30648	7	0,04378	41,82	0,0000
PctWord	0,00165	1	0,00016	0,16	0,7015
MaxNew	0,00005	1	0,00005	0,05	0,8308
PctWor*MaxNew	0,00005	1	0,00005	0,05	0,8308
BALANC	0,30564	1	0,30564	291,95	**0,000**
PctWord*BALANC	0,00046	1	0,00046	0,44	0,5264
MaxNew*BALANC	$2,0408e^{-6}$	1	$2,0408e^{-6}$	0,00	0,9659
PctWord*MaxNew*BALANC	0,0001	1	0,0001	0,01	0,7652
Residual	0,00837	8	0,00105		
Total	0,31485	15	0,02099		

Tabela 6: ANOVA da Métrica *F1-score* dos Experimentos

Origem	SS Parcial	*df*	*MS*	F	Prob >F
Modelo	0,31763	7	0,04537	43,94	0,0000
PctWord	0,00223	1	0,00022	0,22	0,6545
MaxNew	0,00013	1	0,00013	0,12	0,7333
PctWor*MaxNew	0,00016	1	0,00016	0,15	0,7069
BALANC	0,31629	1	0,31629	306,32	**0,000**
PctWord*BALANC	0,000548	1	0,00055	0,53	0,4870
MaxNew*BALANC	$8,2199e^{-6}$	1	$8,2199e^{-6}$	0,01	0,9311
PctWord*MaxNew*BALANC	0,00027	1	0,00027	0,26	0,6218
Residual	0,00826	8	0,00103		
Total	0,32589	15	0,02172		

7 CONCLUSÃO

Neste trabalho foram abordadas algumas das dificuldades que os surdos enfrentam, principalmente em contextos de comunicação e obtenção de informação, além disso, foi apresentado como as tecnologias ajudam a enfrentar essas dificuldades por meio da tradução automática.

Foi também discutido como usuários de algumas dessas tecnologias apontam para a falta da expressão de emoções desses tradutores automáticos. Nesse contexto, foi proposta a utilização de uma rede neural para realizar a análise de emoções em textos, a fim de ser utilizada em conjunto com esses tradutores automático.

Para contornar escassez de base de dados em Português Brasileiro voltadas para esse tipo de tarefa, foram planejados experimentos para encontrar os melhores fatores para se aplicar *data augmentation* numa base de dados disponível. Os experimentos apontaram que a utilização de uma base de dados com *data augmentation* balanceada foi um fator significante para obter os melhores resultado, uma acurácia de 94,29% e um *F1-score* de 94,28%, superando os resultados do trabalho que originou a base de dados, fato que mostra que a estratégia aqui abordada traz uma influência positiva para uma melhora na tarefa de análise de emoções de textos em Português Brasileiro.

Como trabalhos futuros, pretende-se utilizar o melhor modelo de classificação obtido pelo experimentos em conjunto com um tradutor automático para LIBRAS, como a Suíte VLibras[20, 27] que possui código aberto, além de realizar testes com usuários para aferir sobre a qualidade da análise de emoção obtida. Também destaca-se como trabalho futuro a adição de mais exemplos à base de dados para um melhor tratamento de um contexto geral de utilização.

REFERÊNCIAS

[1] Tiago Maritan Ugulino de Araújo. 2012. *Uma solução para geração automática de trilhas em Língua Brasileira de Sinais em conteúdos multimídia.* Ph.D. Dissertation. Universidade Federal do Rio Grande do Norte.

[2] Yunhe Bai and David Bruno. 2020. Addressing Communication Barriers Among Deaf Populations Who Use American Sign Language in Hearing-Centric Social Work Settings. *Columbia Social Work Review* 18, 1 (2020), 37–50.

[3] Vishwanath Bijalwan, Vinay Kumar, Pinki Kumari, and Jordan Pascual. 2014. KNN based machine learning approach for text and document mining. *International Journal of Database Theory and Application* 7, 1 (2014), 61–70.

[4] James Bradbury, Stephen Merity, Caiming Xiong, and Richard Socher. 2016. Quasi-recurrent neural networks. *arXiv preprint arXiv:1611.01576* (2016).

[5] Siddhartha Brahma. 2018. Improved Sentence Modeling using Suffix Bidirectional LSTM. *arXiv preprint arXiv:1805.07340* (2018).

[6] Necati Cihan Camgoz, Simon Hadfield, Oscar Koller, Hermann Ney, and Richard Bowden. 2018. Neural sign language translation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 7784–7793.

[7] Géraldine Coppin and David Sander. 2016. Theoretical approaches to emotion and its measurement. In *Emotion measurement*. Elsevier, 3–30.

[8] Marcele Pereira da Rosa Zucolotto, Luciana Rodrigues Ruiz, and Najara Ferrari Pinheiro. 2019. REFLEXÕES SOBRE LINGUAGEM, SOCIEDADE E SURDEZ. *Revista Uniabeu* 12, 30 (2019), 134–147.

[9] Valeria de Paiva, Alexandre Rademaker, and Gerard de Melo. 2012. OpenWordNet-PT: An Open Brazilian Wordnet for Reasoning. In *Proceedings of COLING 2012: Demonstration Papers*. The COLING 2012 Organizing Committee, Mumbai, India, 353–360. http://www.aclweb.org/anthology/C12-3044 Published also as Techreport http://hdl.handle.net/10438/10274.

[10] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *NAACL-HLT (1)*.

[11] Mariza Miola Dosciatti, LPC Ferreira, and EC Paraiso. 2013. Identificando emoçoes em textos em português do brasil usando máquina de vetores de suporte em soluçao multiclasse. *ENIAC-Encontro Nacional de Inteligência Artificial e Computacional. Fortaleza, Brasil* (2013).

[12] Julian Eisenschlos, Sebastian Ruder, Piotr Czapla, Marcin Kadras, Sylvain Gugger, and Jeremy Howard. 2019. MultiFiT: Efficient Multi-lingual Language Model Fine-tuning. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. 5706–5711.

[13] Hossam Faris, Ibrahim Aljarah, and Seyedali Mirjalili. 2016. Training feedforward neural networks using multi-verse optimizer for binary classification problems. *Applied Intelligence* 45, 2 (2016), 322–332.

[14] Yoav Goldberg. 2017. Neural network methods for natural language processing. *Synthesis Lectures on Human Language Technologies* 10, 1 (2017), 1–309.

[15] Jeremy Howard and Sebastian Ruder. 2018. Universal Language Model Fine-tuning for Text Classification. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 328–339.

[16] Rajkumar S Jagdale, Vishal S Shirsat, and Sachin N Deshmukh. 2016. Sentiment analysis of events from Twitter using open source tool. *IJCSMC* 5, 4 (2016), 475–485.

[17] Rie Johnson and Tong Zhang. 2017. Deep pyramid convolutional neural networks for text categorization. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 562–570.

[18] Sang-Ki Ko, Chang Jo Kim, Hyedong Jung, and Choongsang Cho. 2019. Neural sign language translation based on human keypoint estimation. *Applied Sciences* 9, 13 (2019), 2683.

[19] Jiwei Li, Xinlei Chen, Eduard Hovy, and Dan Jurafsky. 2016. Visualizing and Understanding Neural Models in NLP. *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (2016). https://doi.org/10.18653/v1/n16-1082

[20] Manuella Aschoff Cavalcanti Brandão Lima et al. 2015. *Tradução Automática com Adequação Sintático-semântica para LIBRAS.* Master's thesis. Universidade Federal da Paraíba.

[21] Edward Loper and Steven Bird. 2002. NLTK: the natural language toolkit. *arXiv preprint cs/0205028* (2002).

[22] Marc Moreno Lopez and Jugal Kalita. 2017. Deep Learning applied to NLP. *arXiv preprint arXiv:1703.03091* (2017).

[23] George A Miller. 1995. WordNet: a lexical database for English. *Commun. ACM* 38, 11 (1995), 39–41.

[24] Rodrigo Moraes, JoãO Francisco Valiati, and Wilson P GaviãO Neto. 2013. Document-level sentiment classification: An empirical comparison between SVM and ANN. *Expert Systems with Applications* 40, 2 (2013), 621–633.

[25] Anna Mosolova, Vadim Fomin, and Ivan Bondarenko. 2018. Text Augmentation for Neural Networks.. In *AIST (Supplement)*. 104–109.

[26] Paula M Niedenthal and François Ric. 2017. *Psychology of emotion.* Psychology Press.

[27] Tiago Oliveira, Paula Escudeiro, Nuno Escudeiro, Emanuel Rocha, and Fernando Maciel Barbosa. 2019. Automatic sign language translation to improve communication. In *2019 IEEE Global Engineering Education Conference (EDUCON)*. IEEE, 937–942.

[28] Eslam Omara, Mervat Mosa, and Nabil Ismail. 2019. Emotion Analysis in Arabic Language Applying Transfer Learning. In *2019 15th International Computer Engineering Conference (ICENCO)*. IEEE, 204–209.

[29] Robert Plutchik. 2001. The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American scientist* 89, 4 (2001), 344–350.

[30] Cleomar Rocha and Sarah Caetano de Melgaço. 2018. O uso de aplicativos para tradução de Libras. In *Anais do V Simpósio Internacional de Inovação em Mídias Interativas*.

[31] Hugo Rodrigues and Fábio Libório Rocha. 2016. Uma definição constitutiva de emoções: a constitutive definition of emotions. *Revista Húmus* 5, 15 (2016).

[32] Claire Ryan and Paige Johnson. 2019. Understanding Language Deprivation and Its Role in Deaf Mental Health. *American Annals of the Deaf* 164, 4 (2019), 519–524.

[33] Kashfia Sailunaz, Manmeet Dhaliwal, Jon Rokne, and Reda Alhajj. 2018. Emotion detection from text and speech: a survey. *Social Network Analysis and Mining* 8, 1 (2018), 28.

[34] David Sander. 2013. Models of emotion. *The Cambridge handbook of human affective neuroscience* (2013), 5–56.

[35] Vivek Kumar Singh, Rajesh Piryani, Ashraf Uddin, and Pranav Waila. 2013. Sentiment analysis of movie reviews: A new feature-based heuristic for aspect-level sentiment classification. In *2013 International Mutli-Conference on Automation, Computing, Communication, Control and Compressed Sensing (iMac4s)*. IEEE, 712–717.

[36] Stephanie Stoll, Necati Cihan Camgöz, Simon Hadfield, and Richard Bowden. 2018. Sign Language Production using Neural Machine Translation and Generative Adversarial Networks.. In *BMVC*. 304.

[37] Chi Sun, Xipeng Qiu, Yige Xu, and Xuanjing Huang. 2019. How to fine-tune BERT for text classification?. In *China National Conference on Chinese Computational Linguistics*. Springer, 194–206.

[38] Bayya Yegnanarayana. 2009. *Artificial neural networks.* PHI Learning Pvt. Ltd.