

Uma Proposta de Tradutor Automático com Análise de Emoções de Textos em Português Brasileiro para Libras

Vinícius Matheus Veríssimo da Silva, Rostand Costa

LAVID/UFPB

Centro de Informática – Universidade Federal da Paraíba

{vinicius.matheus, rostand}@lavid.ufpb.br

Resumo: As áreas tecnológicas apresentam alternativas que auxiliam a comunidade surda na aquisição e compartilhamento de informações, como é o caso dos tradutores automáticos. Embora tal abordagem seja bem-vista pela comunidade surda, ela ainda sofre críticas frente a falta de humanidade dessas soluções. Diante disso, esse estudo tem por objetivo realizar a adaptação do tradutor automático de Português Brasileiro para Libras da Suíte VLibras, de modo que o avatar 3D da plataforma performe as emoções contidas na tradução, através de expressões faciais. Apesar de inicial, 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.

Palavras-chave: *acessibilidade; tradução automática; análise de emoções.*

1. Introdução

No Brasil, por volta de 10 milhões de pessoas possuem algum grau de perda auditiva grave (cerca de 5% da população), onde 2,7 milhões possuem um grau de surdez absoluta, conforme censo de 2010 do IBGE. Assim, uma fatia considerável da população brasileira está suscetível à total ou parcial privação de informações, visto que as mesmas são voltadas predominantemente para a população ouvinte.

Uma vez que a população surda tem como língua prioritária a língua de sinais, e estar em um meio onde o que predomina são as línguas orais faz com que isso acarrete resultados negativos em seus processos de desenvolvimento cognitivo e de aprendizado [2, 3], o que impacta diretamente em como uma pessoa surda se integra e participa de sua comunidade [4], uma vez que a sua capacidade de absorver e difundir informações é limitada.

Postas tais dificuldades, várias estratégias aparecem como forma de diminuir as barreiras que as pessoas surdas enfrentam, como a presença de intérpretes humanos em diversos ambientes. Porém, há casos onde a utilização de um intérprete humano não é possível, como na internet. Diante disso, uma solução é a utilização de tradutores automáticos de conteúdo multimídia em língua oral para língua de sinais, como os explorados nos trabalhos de Oliveira et al. [6], Stoll et al. [7] e Araújo [8].

O uso de tradutores automáticos é uma alternativa viável para que a população surda tenha maior acesso a informações, porém, alguns problemas são apontados no estudo de Rocha e Melgaço [5]. Ao realizarem uma pesquisa de opinião com usuários de aplicativos de tradução automática de Português Brasileiro para Libras (Língua Brasileira de Sinais), os autores constaram que há aceitação dessas tecnologias por parte da comunidade surda, mas um ponto amplamente criticado foi a falta de expressões faciais e corporais adequadas no avatar 3D utilizado nessas ferramentas.

Nesse sentido, o objetivo deste trabalho é realizar a adaptação de um tradutor automático de língua oral para

língua de sinais, visando possibilitar a expressão de emoções no avatar 3D, por meio de expressões faciais no tradutor da Suíte VLibras [8].

O restante do trabalho está organizado da seguinte forma: na Seção 2 será apresentada a metodologia de adaptação do tradutor automático; na Seção 3 serão apresentados os resultados dessa adaptação; na Seção 4 será realizada uma discussão acerca do trabalho apresentado; e, por fim, a Seção 5 trará a conclusão do trabalho.

2. Metodologia

O processo de adaptação do tradutor automático será realizado sobre o tradutor automático da Suíte VLibras. A Suíte VLibras é um conjunto de ferramentas para a tradução de conteúdo multimídia de Português Brasileiro para Libras, sinalizada por um avatar 3D. Esse sistema é composto de programas para computadores pessoais, aplicativos móveis, *plug-in web*, serviços de nuvem, entre outros. Além disso, ele pode ser aplicado em diversas esferas, sendo ele atualmente utilizado até mesmo por alguns sites públicos governamentais.

De forma mais específica, ressalta-se que a adaptação do tradutor automático será realizada no serviço de tradução em nuvem e na versão de *desktop* do sinalizador automático da Suíte VLibras. Processo que será descrito nas próximas seções.

2.1 Serviço de Tradução com Emoção

Para a criação do serviço de tradução com emoção, será utilizado o modelo de análise de emoção em textos Português Brasileiro proveniente do trabalho de Veríssimo e Costa [9]. O presente estudo utilizará o modelo que obteve melhores resultados computacionais. No trabalho de Veríssimo e Costa [9], os autores exploraram o aumento de uma base de dados de manchetes jornalísticas, e a criação de uma rede neural capaz de fazer a análise de emoções de textos em Português Brasileiro. As emoções compreendidas pelo

estudo são: alegria, raiva, tristeza, medo, nojo, surpresa e a ausência de emoção (neutra).

No que se refere ao tradutor automático, o serviço de tradução original da Suíte VLibras realiza a tradução de texto em Português Brasileiro para glosa Libras, uma forma de notação escrita das Línguas de Sinais. No que tange à adaptação desse serviço, será na forma de um *middleware*, que realizará as requisições de tradução para o serviço original e para o analisador de emoções simultaneamente.

Nesse serviço de tradução adaptada, o analisador de emoções e o tradutor automático original estão sendo executados em paralelo e em ambientes diferentes. O mesmo texto de entrada será enviado para ambos os algoritmos, ou seja, a análise de emoções e tradução serão realizadas sobre o mesmo texto em Português Brasileiro. Essa simultaneidade é positiva, pois a informação do texto não será modificada, sendo modificada apenas a sua forma de transmissão.

A Figura 1 mostra um exemplo genérico do fluxo de execução do serviço de tradução adaptado. É possível visualizar três módulos: o módulo de Classificação de Emoções, responsável por realizar a análise de emoções do texto; o módulo de Tradução, responsável por realizar a tradução do texto em glosa; e o Tradutor com Emoções, que se utiliza do resultado dos outros módulos para criar uma glosa com a informação de emoção.

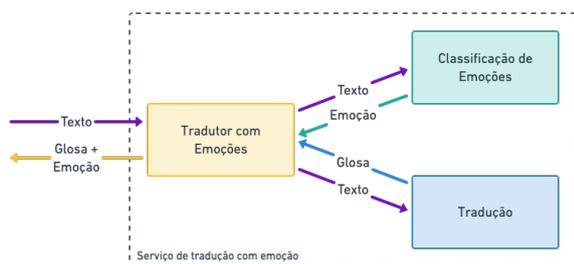


Figura 1. Exemplo genérico do fluxo do serviço de tradução adaptado.

Para a adição da emoção da glosa foi utilizada uma abordagem baseada em marcadores, como os utilizados nas linguagens de marcação HTML e XML. Os marcadores contendo os caracteres < e > indicam o início da emoção na porção do texto, e os marcadores contendo os caracteres <\ e >, denotam o fim da emoção.

A Figura 2 apresenta um exemplo de utilização dos marcadores na glosa. O exemplo “a)” é uma forma simples de utilização, com apenas uma emoção para todo o texto. Por outro lado, o exemplo “b)” apresenta uma forma mais complexa de utilização, com múltiplos marcadores, trazendo assim, múltiplas emoções em um mesmo texto.

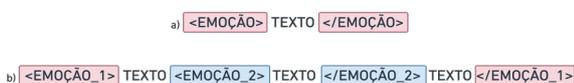


Figura 2. Exemplo de utilização dos marcadores de emoções na glosa

Para essa adaptação, apenas o exemplo a) da Figura 2 será abordado, uma vez que o analisador de emoções de Veríssimo e Costa [9] realiza a identificação de uma emoção por texto.

O serviço de tradução adaptado foi criado utilizando a linguagem Python, e o *micro-framework* Flask para prover uma API de acesso. Além disso, foram utilizadas bibliotecas para auxiliar no paralelismo das operações, e o *framework* fastai, para o uso dos modelos de análise de emoções. O repositório do projeto pode ser acessado por esse [link](#).

2.2 Sinalizador Automático com Emoções

A adaptação do sinalizador automático *desktop* da Suíte VLibras necessitará de modificação no código-fonte, que foi cedido pelo LAVID (Laboratório de Vídeo Digital), o qual desenvolve o projeto.

O sinalizador automático funciona realizando uma requisição ao serviço de tradução automática para que seja feita a tradução do texto em Português Brasileiro para glosa Libras. Para o sinalizador, cada palavra da glosa se refere a um *bundle* de animação, que descreve como o avatar 3D deve performar o sinal. Além disso, uma legenda acompanha a sinalização sincronizadamente.

A adaptação do sinalizador automático deverá interpretar a nova glosa fornecida pelo serviço de tradução adaptado que foi apresentado na Seção 2.1, de modo que ele deve apresentar a emoção contida na nova glosa, por meio das expressões faciais do avatar 3D, durante a sinalização das demais palavras da glosa. Além disso, essa informação de emoção não será apresentada na legenda que acompanha a sinalização, sendo apresentada apenas no avatar.

Em relação à definição das expressões faciais do avatar 3D, para cada uma das emoções abordadas, foi utilizado como base o estudo de Silva et al. [10], que descreve as expressões faciais da Libras na notação do sistema FACS, primariamente definido por Ekman e Friesen [11].

No FACS, as expressões faciais são definidas em termos de unidade de ação (AUs - *action units*), 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).

A Tabela 1 traz uma adaptação da tabela apresentada no estudo de Silva et al. [10], que utilizou os valores definidos pela literatura para a expressões faciais das emoções, argumentando que as expressões faciais de emoções são amplamente utilizadas e compreendidas pela comunidade surda.

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 da pálpebra superior; AU6 ao levantador da bochecha; AU7 ao apertador de pálpebra; AU9 ao enrugador de nariz; AU12 ao puxador de canto dos lábios; AU15 ao

depressor de canto dos lábios; *AU16* ao depressor do lábio inferior; *AU20* ao esticador de lábios; *AU23* ao endurecedor de lábio; e *AU26* ao atuador responsável pela queda do queixo.

Tabela 1. Emoções básicas na notação de FACS, adaptada de [10].

FACS	Descrição
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	Nojo
AU9 + AU15 + AU16	

Para a reprodução dessas descrições das expressões faciais das emoções básicas no avatar 3D, foram utilizadas as animações de expressões faciais preestabelecidas no projeto, apesar das mesmas não possuírem paralelos exatos com todas as AUs, são elas: sorriso, fechar sobrancelhas, abrir sobrancelhas, abaixar sobrancelhas, subir sobrancelhas, franzir sobrancelhas, subir lábio superior, contrair lábios, baixar cantos da boca, subir cantos da boca, inflar bochechas, fazer bico, e contrair bochechas.

Em relação ao expressar das emoções, para expressar Alegria, foram modificados os valores das animações de sorriso e cantos da boca. Para expressar Tristeza, foram modificados os valores das animações das sobrancelhas, cantos da boca e contrair lábio. Para expressar Raiva, foram modificados os valores das animações das sobrancelhas, fazer bico e contrair os lábios. Para expressar Medo, foram modificadas as animações das sobrancelhas e bochechas, e também foi utilizada a animação de sorriso, sendo corrigido modificando o lábio superior e cantos da boca. Para expressar Nojo, foram modificados os valores das animações dos cantos da boca, contrair lábios, e também a animação de franzir as sobrancelhas, visando poder enrugando um pouco o nariz; sendo corrigida com a animação de abrir as sobrancelhas. Por fim, para expressar Surpresa, foram modificados os valores das animações das sobrancelhas, subir lábio superior. A expressão padrão do avatar 3D é considerada como a expressão do sentimento Neutro.

3. Resultados

A adaptação do tradutor e sinalizador automático da Suíte VLibras possibilitou a representação das emoções básicas, sendo elas: alegria, tristeza, raiva, medo, nojo e surpresa. Além da emoção neutra, a qual é performada pela ferramenta de forma padrão. Na Figura 3 é apresentado um *frame* da sinalização da palavra *ANDAR* nas 6 emoções básicas e na emoção neutra.

Para uma melhor avaliação da adaptação, também foi produzido um vídeo demonstrando a sinalização com emoção de textos contidos na base de dados de testes do analisador de emoções do estudo de Veríssimo e Costa [9]. O referido vídeo pode ser acessado por esse [link](#).

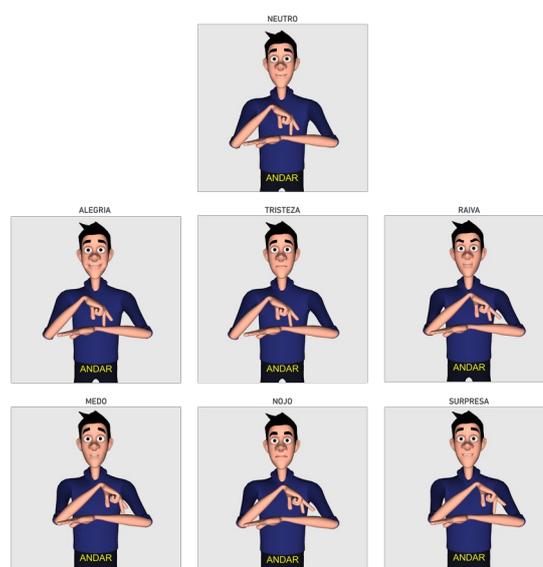


Figura 3. Exemplo da sinalização da palavra *ANDAR* na emoção neutra, alegria, tristeza, raiva, medo, nojo e surpresa.

Uma dificuldade encontrada pelo estudo foi que alguns movimentos importantes, como abrir a boca e controle da pálpebra não foram possíveis, frente às animações pré-definidas do avatar, o que prejudicou a definição de algumas expressões, como o medo e a surpresa. Além disso, ressalta-se que os testes foram realizados apenas com o avatar masculino, visto que, apesar de a ferramenta conter um avatar feminino, ele não apresenta as mesmas animações pré-definidas para expressões faciais, fazendo com que elas não sejam tratadas da mesma forma.

4. Discussão

A adaptação do tradutor automático de textos em Língua Oral para Língua de Sinais realizada a partir do fluxo genérico apresentado, e da sintaxe de construção da glosa com emoções, serve como um ponto inicial para as inúmeras possibilidades de implementação de um tradutor adaptado. A adaptação Suíte VLibras mostrou como a adaptação do tradutor automático pode ser feita com pouco impacto, mas em relação ao sinalizador, é necessária uma abordagem mais profunda.

Um ponto a ser mais trabalhado futuramente é a sintaxe de adição de emoções à glosa, que pode ser ampliada com a adição de metadados, ou atributos aos marcadores 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.

5. Conclusões

Em um cenário onde uma parcela significativa da população, sendo essa a população de pessoas surdas, vivem 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.

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, podendo resolver alguns dos principais problemas referentes à expressividade dos tradutores automáticos.

Bibliografia

- [1] Dizeu, L. C. T. D. B.; Caporali, S. A. (2005) A língua de sinais constituindo o surdo como sujeito. *Educação & Sociedade* 26(91): 583-597.
- [2] Bai, Y.; Bruno, D. (2020) 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. DOI: [10.7916/cswr.v18i1.5928](https://doi.org/10.7916/cswr.v18i1.5928)
- [3] Ryan, C.; Johnson, P. (2019) Understanding language deprivation and its role in deaf mental health. *American Annals of the Deaf* 164(4): 519-524. DOI: [10.1353/aad.2019.0030](https://doi.org/10.1353/aad.2019.0030)
- [4] Zucolotto, M.; Ruiz, L.; Pinheiro, N. (2019) Reflexões Sobre Linguagem, Sociedade e Surdez. *Revista Uniabeu* 12(30): 134-147.
- [5] Rocha, C.; Melgaço, S. (2018) O uso de aplicativos para tradução de Libras. *Anais do V Simpósio Internacional de Inovação em Mídias Interativas*, 36-47.
- [6] Oliveira, T.; Escudeiro, P.; Escudeiro, N.; Rocha, E.; Barbosa, F. (2019) Automatic sign language translation to improve communication. 2019 IEEE Global Engineering Education Conference (EDUCON), 937-942. DOI: [10.1109/EDUCON.2019.8725244](https://doi.org/10.1109/EDUCON.2019.8725244)
- [7] Stoll, S.; Camgöz, N.; Hadfield, S.; Bowden, R. (2018) Sign language production using neural machine translation and generative adversarial networks. In: *Proceedings of the 29th British Machine Vision Conference (BMVC 2018)*. British Machine Vision Association.
- [8] Araújo, T. (2012) Uma solução para geração automática de trilhas em Língua Brasileira de Sinais em conteúdos multimídia. Tese. Programa de Pós-graduação em Engenharia Elétrica e da Computação. https://repositorio.ufrn.br/jspui/bitstream/123456789/1519/0/1/TiagoMUA_TESE.pdf
- [9] Veríssimo, V.; Costa, R. (2020) Using Data Augmentation and Neural Networks to Improve the Emotion Analysis of Brazilian Portuguese Texts. In: *Proceedings of the Brazilian Symposium on Multimedia and the Web*, p. 13-20. DOI: [10.1145/3428658.3431080](https://doi.org/10.1145/3428658.3431080)
- [10] Silva, E.; Costa, P.; Kumada, K.; Martino, J.; Florentino, G. (2020) Recognition of affective and grammatical facial expressions: a study for Brazilian sign language. *European Conference on Computer Vision*, Springer, Cham, 218-236. DOI: [10.1007/978-3-030-66096-3_16](https://doi.org/10.1007/978-3-030-66096-3_16)
- [11] Eckman, P., & Friesen, W. (1977) *Manual for the facial action coding system*. Consulting Psych. Press, Palo Alto.