

# **TRALP - Tradução Libras Português Implementada por Software em Dispositivos Móveis**

Marcelo Krieger  
Vanessa Palomo

**Internal Research Reports**

Number 50 | September 2016

# **TRALP - Tradução Libras Português Implementada por Software em Dispositivos Móveis**

Marcelo Krieger  
Vanessa Palomo

## **CREDITS**

### **Publisher:**

**MAXWELL / LAMBDA/CCPA/VRAC**

**Sistema Maxwell / Laboratório de Automação de Museus, Bibliotecas Digitais e Arquivos**

**<http://www.maxwell.vrac.puc-rio.br/>**

### **Organizers:**

Alexandre Street de Aguiar

Delberis Araújo Lima

### **Cover:**

Ana Cristina Costa Ribeiro

# TRALP – Tradução Libras Português implementada por software em dispositivos móveis

Marcelo Krieger, Vanessa Palomo

LAMBDA – Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio)  
Caixa Postal 38.097 – 22.451-900 – Rio de Janeiro – RJ – Brazil

mkrieger@esp.puc-rio.br, vpalomo@puc-rio.br

**Abstract.** *The TRALP Project aims to build an entirely software implemented Libras (Brazilian Sign Language) – Portuguese and Portuguese – Libras translator for portable devices. TRALP will use the device's built-in camera to identify Libras visual-spatial and kinetic-visual articulation. Microphone and keyboard will voice and text in Portuguese. The input thus obtained will be morphologically and syntactically evaluated for the input language and transformed to the corresponding message in the output language. Loudspeaker and screen will show Portuguese output and an Avatar will gesture the output in Libras. This paper describes the main concepts, challenges and solutions found during TRALP development.*

**Resumo.** *O TRALP consiste em um TRAdutor Libras Português e Português – Libras inteiramente implementado através de software para dispositivos móveis. O TRALP utilizará a câmera do dispositivo para identificar a articulação visual-espacial e cinésico-visual de Libras. Voz e texto em português serão capturados por microfone e teclado. A entrada obtida será submetida às análises morfológica e sintática do idioma de entrada e transformada na sua correspondente no idioma de saída. O alto-falante e tela serão usados para exibir a saída em português e um Avatar gesticulará a saída em Libras. Neste artigo são descritos os principais conceitos do TRALP, os desafios encontrados e as soluções para resolvê-los.*

## 1. Introdução

**Libras** é a língua de sinais usada pela maioria dos surdos dos centros urbanos brasileiros. É derivada tanto de uma língua de sinais autóctone quanto da língua gestual francesa. Por isso, é semelhante a outras línguas de sinais da Europa e da América. Libras não é a simples gestualização da língua portuguesa, e sim uma língua à parte, como é comprovado pelo fato de que em Portugal usa-se uma língua de sinais diferente, a Língua Gestual Portuguesa (LGP).

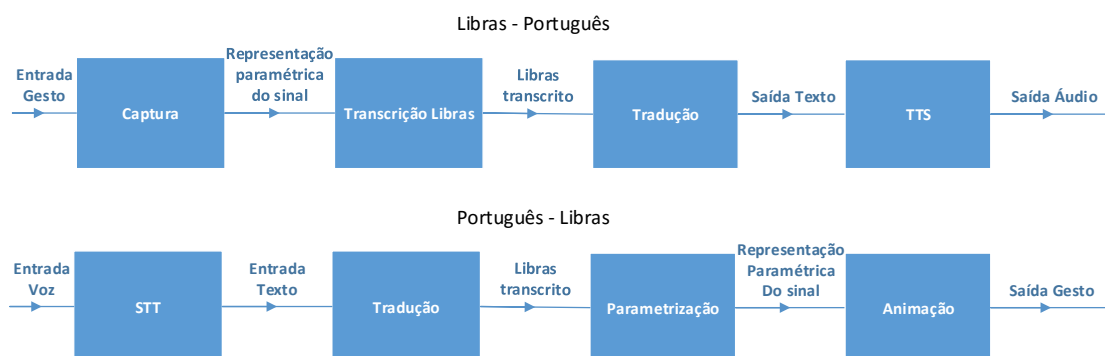
De acordo com Sylvia Lia Grespan Neves, professora da disciplina de Libras da Faculdade de Ciências Médicas da Santa Casa de São Paulo, e que também é surda, ainda existe uma grande dificuldade para os não ouvintes no Brasil, principalmente pela falta de preparo de profissionais nos serviços básicos, que desconhecem totalmente a língua de sinais. [Neves, S. 2013]. É neste contexto que torna-se relevante o desenvolvimento de ferramentas que permitam a comunicação entre pessoas que, de fato, comunicam-se em diferentes idiomas.

Acrescente-se a isto, a explosão na adoção de dispositivos portáteis “inteligentes”. Tomando os smartphones como exemplo. Segundo a consultoria IDC, foram vendidos 54,5 milhões de smartphones no ano de 2014 no Brasil. [IDC Releases 6/4/2015]. A junção destes dois fatores justifica o desenvolvimento de mecanismos de auxílio à comunicação entre as comunidades surda e ouvinte usando dispositivos móveis.

O TRALP propõe-se a usar somente os recursos padrões existentes no dispositivo para atingir a finalidade proposta. Esta abordagem difere de outras iniciativas neste campo que têm usado dispositivos auxiliares externos, como luvas (datagloves) para obter a entrada de dados sinalizada. Estas opções tornam proibitiva e/ou pouco prática a adoção da aplicação.

## 2. Estrutura geral do Aplicativo

Os diagramas de blocos abaixo representam o processo pelos quais uma mensagem deve passar para ser traduzida para o outro idioma.



**Figura 1. Estrutura conceitual dos processos do TRALP**

Na tradução Libras → Português, o passo inicial consiste em capturar a gesticulação através da câmera do dispositivo. Em seguida, a sequência de imagens obtida é processada para chegar à transcrição de Libras, como a encontrada em [Primecurso]. Submete-se a mensagem em Libras transcrito ao processo de análise e transformação para obter o texto da mensagem em português. Esta mensagem é exibida na tela do dispositivo e pode-se solicitar que a mensagem seja ouvida através do alto-falante (TTS).

Na tradução Português → Libras, o primeiro passo é a captura da mensagem que pode ser texto ou voz. Neste último caso, deverá haver um processo de conversão de voz para texto (STT), antes de prosseguir. O passo seguinte é a tradução da mensagem de texto em português para a transcrição de Libras. Em seguida a transcrição obtida é parametrizada para que finalmente seja gesticulada em tela por um Avatar.

## 3. Problemas e soluções encontrados

Algumas premissas básicas foram seguidas no projeto do TRALP:

- Ser implementado totalmente em software. Esta foi a premissa fundamental do Projeto. O desenvolvimento cobre as três plataformas principais: iOS (hardware e software da Apple), Android (desenvolvida pelo Google e adotada por diversos

fabricantes) e Windows Phone (desenvolvida pela Microsoft e adotada principalmente por Nokia e HTC).

- Ser, no máximo possível, independente da Internet. Tanto o vocabulário quanto os algoritmos devem caber no espaço de memória disponível no dispositivo.

Além destas linhas mestras, algumas questões específicas precisaram ser tratadas em cada um dos módulos componentes do TRALP.

### 3.1. Captura e tratamento dos gestos

A captura dos gestos inicia com o correto enquadramento da pessoa que irá gesticular. O conceito de Realidade Aumentada foi usado para auxiliar o enquadramento. Ao acionar a câmera, é exibida uma silhueta à qual deve ser “encaixada” a pessoa que irá gesticular. A silhueta é exibida em vermelho enquanto o correto enquadramento não é obtido, e muda para verde quando a pessoa está na posição certa. Para se chegar ao correto enquadramento foram estudadas duas alternativas: (1) usar a API de identificação de rostos quando existir na plataforma e comparar a posição do rosto com a posição da silhueta ou; (2) aplicar um operador de localização de bordas [Hueckel, 1971] para comparar a localização da borda do objeto com a silhueta. (Figura 2).



**Figura 2. Enquadramento – Eduardo Carvalho, bolsista do Projeto**

Uma vez iniciada a captura das imagens, a decisão seguinte passa a ser a de selecionar o *frame rate* adequado. Um excesso de quadros trará informação redundante e resultará em um tempo de processamento excessivo, enquanto a falta de quadros poderá causar perda na mensagem. A hipótese básica é a de que o gesticulador buscará ser compreendido pelo seu interlocutor. Seguindo esta hipótese, a execução natural dos gestos permite a adoção de um frame rate de 10 fps.

As imagens capturadas são usadas para identificar os gestos realizados. O processamento da imagem identificará os elementos componentes de um gesto em Libras: configuração das mãos, ponto de articulação, orientação de palma, movimento e expressão facial.

O espaço disponível não é suficiente para uma descrição completa das técnicas e algoritmos usados, mas resumidamente exemplificando com a posição da mão: o espaço ocupado pelo gesticulador foi dividido em regiões, nas três dimensões. Foram identificadas 7 posições no eixo horizontal, 14 no eixo vertical. O sistema busca a mão, através da mesma técnica usada para o enquadramento do gesticulador. Para a profundidade, basta identificar se a mão está em contato com o corpo ou adiante deste, o que é calculado através da proporção entre o tamanho da mão em relação ao tamanho do gesticulador. A posição é expressa como um dos parâmetros da LPS (ver na próxima seção).

### 3.2. Linguagem Paramétrica de Sinais (LPS)

O Português é formado através da combinação das 26 letras e alguns poucos modificadores (acentos, por exemplo) e sinais gráficos. Para armazenar cada uma destas possibilidades, os computadores as representam através de uma combinação de oito “0”s e “1”s, chamada Byte. Desta forma, o armazenamento de textos extensos pode ser feito de maneira bastante econômica. As obras completas de Shakespeare exigem somente 5 Megabytes (5 milhões de caracteres), por exemplo.

Mas as palavras em Libras são expressas através de imagens, como na Figura 3.

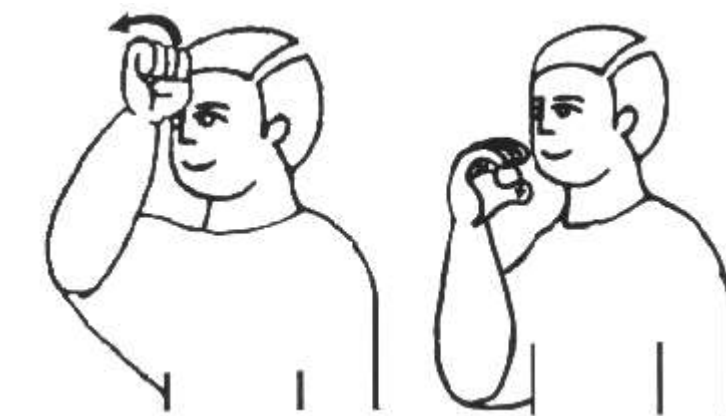


Figura 3. “Pensar” e “Sábado” em Libras

São necessários 100 Kilobytes para armazenar 1 imagem em baixa resolução em um computador. Um vocabulário de 3.000 sinais exigiria 300 Megabytes, isto se cada sinal em Libras pudesse ser guardado em apenas uma imagem. Não é o caso. Muitas palavras em Libras incluem movimento em sua formação. Precisariamos armazenar, portanto, vídeos. O que faz crescer a estimativa da necessidade de armazenamento dos mesmos 3.000 sinais para 1,5 Gigabytes. Este número ocuparia toda a memória de muitos dispositivos, além de exigir muita banda de conexão para instalação do aplicativo. Sem contar, o tempo de processamento necessário para comparar as imagens capturadas com as que compõem o vocabulário.

Para resolver esta questão, foi feito o mapeamento dos parâmetros componentes de Libras, descritos ao final da seção anterior, formando uma Linguagem Paramétrica de Sinais (LPS). Foram identificadas 74 configurações de mão e 82 pontos de articulação. Logo, bastam dois Bytes para armazenar cada uma das possíveis combinações de pares. São necessários mais itens para os demais parâmetros. Ainda assim, ao final, usando a

LPS chegou-se a uma necessidade de apenas 15 Megabytes para armazenar todo o vocabulário inicial. A palavra “pensar”, por exemplo, é articulada na testa (posição de mão 100), sendo esta a única diferença que tem com a palavra “sábado”, que é articulada no queixo (posição de mão 108).

### 3.3. Conversão voz x texto (STT e TTS)

Esta necessidade foi inteiramente resolvida através das APIs disponíveis nas plataformas. Os únicos compromissos que tiveram que ser feitos foram:

- Na plataforma Android, há a necessidade de instalar um módulo específico para implementar o recurso.
- Na plataforma iOS, o recurso está disponível somente se o dispositivo está conectado à Internet.

### 3.4. Conversão de idioma

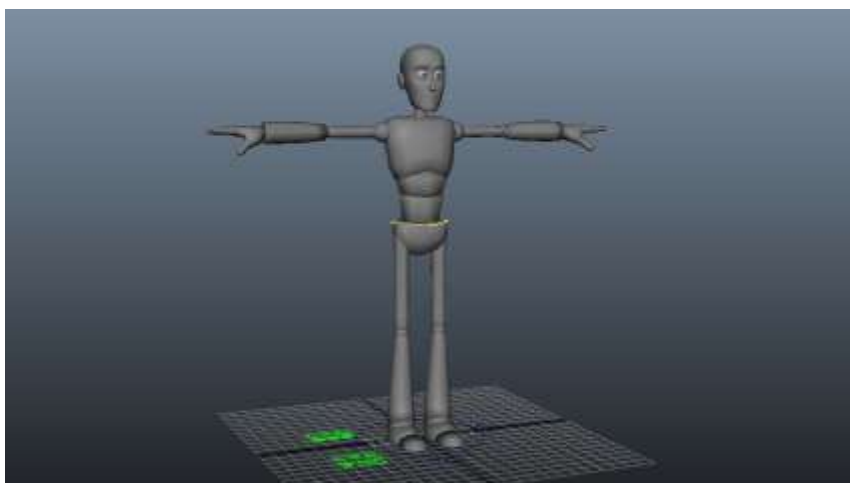
Esta parte do Projeto realiza a tradução de um texto em Português para um texto em Libras transcrito e vice-versa. As seguintes premissas foram seguidas:

- O sistema deve realizar a tradução de idiomas, seguindo a estrutura sintática e semântica de cada um.
- Se, na tradução para Libras, a palavra não estiver no dicionário, à mesma é representada por sua datilologia (sinalização letra a letra).
- Para atingir a finalidade proposta, o sistema deverá realizar as seguintes atividades:
  - Análise Morfológica - análise das palavras para classificação nas classes gramaticais existentes.
  - Análise Sintática - análise da frase para classificação em sujeito, verbo e objeto.
  - Transformação - modificações na frase para que esta seja traduzida por inteiro, respeitando a gramática de ambas as línguas e não palavra a palavra.
  - Contextualização – desambiguação das sentenças, para que a tradução seja mais precisa, e leve em conta o contexto da sentença.
  - Tradução palavra a palavra – feita através de um dicionário, que para cada palavra em uma língua, possui o correspondente na outra língua.

### 3.5. Animação de Avatar

O passo final na conversão Português → Libras é a exibição dos gestos na tela do dispositivo. Isto é feito através de um Avatar. O termo designa as figuras que são criadas à imagem do usuário, permitindo sua "personalização" no interior das máquinas e telas de computador.

O processo envolve duas etapas: construção do Avatar e animação do Avatar. Dentre as várias técnicas disponíveis para a modelagem do Avatar, foi escolhida a denominada “*Skeletal Subspace Transformation*”, que envolve a construção de uma estrutura similar a um esqueleto sob o modelo 3D. Este processo é conhecido como “rigging”. Foi selecionado um personagem denominado Norman Rig (Figura 4), disponível à comunidade de animadores, como ponto de partida para o Avatar.



**Figura 4. Norman Rig**

A ferramenta Blender (<http://www.blender.org/>) foi escolhida para executar a animação do Avatar através da passagem dos parâmetros da LPS, por permitir a integração dentro das plataformas desejadas.

#### **4. Estágio de desenvolvimento**

No momento atual, junho de 2015, os seguintes itens estão totalmente concluídos: (1) Especificação do Projeto, (2) Interface com o usuário, (3) Conversão voz x texto e (4) Linguagem Paramétrica de Sinais. Este último permite que o vocabulário do projeto esteja sendo estruturado e carregado. Os demais itens: (1) Captura e Tratamento dos Gestos, (2) Conversão de idioma e (3) Animação de Avatar estão em processo de codificação e testes. A expectativa é ter o Projeto concluído até o final do ano corrente.

#### **Agradecimentos**

Este trabalho foi desenvolvido com recursos do CNPq, oriundos da chamada MCTI-SECIS/CNPq nº 84/2013 – Tecnologia Assistiva.

#### **References**

Hueckel, M.H. “An Operator which Locates Edges in Digitized Pictures”, Journal of the Association of Computing Machinery, Vol. 18, No. 1, January 1971, pp. 113.125.

IDC Releases (6/4/2015) “Estudo da IBC Brasil aponta que, em 2014, brasileiros compraram cerca de 104 smartphones por minuto”, <http://br.idclatin.com/releases/news.aspx?id=1801>.

Neves, S. (2013) “Ausência de conhecimento em Libras limita acesso de surdos a serviços básicos da saúde”, Boletim Conectar, edição 28 da Santa Casa de Misericórdia de São Paulo.

Prime Cursos. <https://www.primecursos.com.br/>. Curso de Libras Básico, capítulo 8.