

UNIVERSIDADE REGIONAL DE BLUMENAU
CENTRO DE CIÊNCIAS EXATAS E NATURAIS
CURSO DE CIÊNCIA DA COMPUTAÇÃO – BACHARELADO

PROTÓTIPO DE UM TRADUTOR DE TEXTOS DE LÍNGUA
PORTUGUESA PARA LIBRAS

JONATHAN RODRIGUES EV

BLUMENAU
2015

2015/1-18

JONATHAN RODRIGUES EV

**PROTÓTIPO DE UM TRADUTOR DE TEXTOS DE LÍNGUA
PORTUGUESA PARA LIBRAS**

Trabalho de Conclusão de Curso apresentado ao curso de graduação em Ciência da Computação do Centro de Ciências Exatas e Naturais da Universidade Regional de Blumenau como requisito parcial para a obtenção do grau de Bacharel em Ciência da Computação.

Profa. Joyce Martins, Mestre – Orientadora

**BLUMENAU
2015**

2015/1-18

PROTÓTIPO DE UM TRADUTOR DE TEXTOS DE LÍNGUA PORTUGUESA PARA LIBRAS

Por

JONATHAN RODRIGUES EV

Trabalho de Conclusão de Curso aprovado
para obtenção dos créditos na disciplina de
Trabalho de Conclusão de Curso II pela banca
examinadora formada por:

Presidente: _____
Profa. Joyce Martins, Mestre – Orientadora, FURB

Membro: _____
Prof. Aurélio Faustino Hoppe, Mestre – FURB

Membro: _____
Prof. Dalton Solano dos Reis, Mestre – FURB

Blumenau, 06 de julho de 2015

Dedico este trabalho a todos que de alguma
maneira farão bom uso deste.

AGRADECIMENTOS

Primeiramente, a minha orientadora, Joyce Martins. Sem o seu apoio este trabalho não seria o mesmo.

Ao professor Alexander Roberto Valdameri, por me apresentar este tema.

À professora Fabiana Schmitt Corrêa, pela atenção e ajuda com LIBRAS.

Ao Eric Boeing, pela ajuda na construção do *abstract*.

Aos meus pais, Jorge e Marisete, pelo apoio durante toda essa jornada.

A todos os professores e colegas que de alguma maneira contribuíram para meu aprendizado.

Por fim, à equipe do Google. Não consigo imaginar a conclusão deste curso sem utilizar suas tecnologias.

É uma falta de responsabilidade esperar que alguém faça as coisas por nós.

John Lennon

RESUMO

Este trabalho apresenta a especificação e a implementação de um protótipo para um tradutor de Língua Portuguesa para Língua BRAsileira de Sinais (LIBRAS). O usuário deverá informar textos em Língua Portuguesa para o protótipo realizar o processo de tradução. Inicialmente, é feita a etiquetagem das palavras do texto em português, para então traduzir o texto seguindo a gramática de LIBRAS. Após a tradução ser concluída, o usuário poderá reproduzir vídeos com os sinais em LIBRAS equivalentes ao texto informado. Tais sinais são retirados do Dicionário da Língua Brasileira de Sinais. Nos testes realizados, o protótipo apresentou resultados satisfatórios, mesmo que com limitações, sobretudo em relação à quantidade de sinais presentes no Dicionário da Língua Brasileira de Sinais, obrigando que muitas palavras sejam soletradas em LIBRAS, ainda que implementada uma solução para amenizar essa situação, através da criação de um dicionário de exceções. Além disso, nota-se que é possível refinar o processo de tradução, para atingir resultados com melhor precisão. O protótipo foi implementado na linguagem de programação Java, contando com as bibliotecas auxiliares Cogroo e Xuggle.

Palavras-chave: Processamento de linguagem natural. LIBRAS. Tradução automática. Língua portuguesa.

ABSTRACT

This work presents the specification and implementation of a prototype for a Portuguese-speaking translator for Brazilian Sign Language (LIBRAS). The user must inform texts in Portuguese for the prototype perform the translation process. Initially, it is made in the labeling of Portuguese words of text, and then translate the text following the LIBRAS grammar. After the translation is complete, the user can play videos with the signs in LIBRAS equivalent to the supplied text. Such signals are taken from the Dictionary of Brazilian Sign Language. In tests, the prototype showed satisfactory results, albeit with limitations, especially with regard to the amount of signals present in the Dictionary of Brazilian Sign Language, forcing many words that are spelled in LIBRAS, although a solution implemented to mitigate this situation, by creating an exceptions dictionary. Furthermore, it notes that it is possible to refine the process of translation, to achieve results with better accuracy. The prototype was implemented in the Java programming language, with auxiliary libraries Cogroo and Xuggle.

Key-words: Natural language processing. LIBRAS. Machine translation. Portuguese language.

LISTA DE FIGURAS

Figura 1 - Graduação de LIBRAS - nível de frases.....	17
Figura 2 - Versão <i>online</i> do Dicionário da Língua Brasileira de Sinais.....	18
Figura 3 - ProDeaf para dispositivos móveis	20
Figura 4 - Saída do Poli-Libras.....	21
Figura 5 - Interface do LIBROL.....	23
Figura 6 - Casos de uso	25
Figura 7 - Diagrama de classes.....	28
Figura 8 - Diagrama de sequência	29
Figura 9 - Saída da execução do trecho de código mostrado no Quadro 10	30
Figura 10 - Estrutura do dicionário de exceções	32
Figura 11 - Interface gráfica da aplicação	35
Figura 12 – Execução da aplicação	36

LISTA DE QUADROS

Quadro 1 - Etiquetagem da frase “Conversei com arquiteto para fazer uma casa moderna.”..	19
Quadro 2 - Caso de uso: Abrir arquivo de texto	25
Quadro 3 - Caso de uso: Traduzir para LIBRAS	26
Quadro 4 - Caso de uso: Salvar arquivo de texto	26
Quadro 5 - Caso de uso: Limpar tradução anterior	26
Quadro 6 - Caso de uso: Reproduzir vídeos	27
Quadro 7 - Caso de uso: Pausar e continuar reprodução	27
Quadro 8 - Caso de uso: Mudar a sentença que está sendo exibida	27
Quadro 9 - Caso de uso: Reproduzir a sentença do início	27
Quadro 10 - Trecho de código mostrando a utilização de Cogroo	30
Quadro 11 - Exemplo de regras do dicionário de exceções	32
Quadro 12 - Trecho de código onde são descartados <i>tokens</i> não utilizados em LIBRAS	33
Quadro 13 - Trecho de código que realiza a inserção do sujeito na sentença	34
Quadro 14 – Exemplo de frases em português com a palavra <i>andar</i>	37
Quadro 15 - Comparativo entre trabalhos correlatos e o protótipo desenvolvido.....	38

LISTA DE ABREVIATURAS E SIGLAS

API – *Application Programming Interface*

FENEIS – Federação Nacional de Educação e Integração dos Surdos

FURB – Universidade Regional de Blumenau

IBGE - Instituto Brasileiro de Geografia e Estatística

IDE – *Integrated Development Environment*

INES – Instituto Nacional de Educação de Surdos

LIBRAS – Língua BRAsileira de Sinais

PLN – Processamento de Linguagem Natural

RF – Requisito Funcional

RNF – Requisito Não-Funcional

SVO – Sujeito-Verbo-Objeto

UML – *Unified Modeling Language*

USP – Universidade de São Paulo

XML – *Extensible Markup Language*

SUMÁRIO

1 INTRODUÇÃO.....	12
1.1 OBJETIVOS.....	13
1.2 ESTRUTURA.....	13
2 FUNDAMENTAÇÃO TEÓRICA	14
2.1 LIBRAS.....	14
2.1.1 Sistema de transcrição.....	14
2.1.2 Classes gramaticais	15
2.1.3 Sintaxe.....	16
2.2 DICIONÁRIO DA LÍNGUA BRASILEIRA DE SINAIS	17
2.3 PROCESSAMENTO DE LINGUAGEM NATURAL.....	19
2.4 TRABALHOS CORRELATOS.....	20
2.4.1 PRODEAF	20
2.4.2 POLI-LIBRAS	21
2.4.3 LIBROL	22
3 DESENVOLVIMENTO DO PROTÓTIPO.....	24
3.1 REQUISITOS PRINCIPAIS DO PROBLEMA A SER TRABALHADO.....	24
3.2 ESPECIFICAÇÃO	24
3.2.1 Diagrama de casos de uso	24
3.2.2 Diagrama de classes	27
3.2.3 Diagrama de sequência	29
3.3 IMPLEMENTAÇÃO	29
3.3.1 Técnicas e ferramentas utilizadas.....	30
3.3.2 Tradução de sentenças de Língua Portuguesa para LIBRAS.....	31
3.3.3 Operacionalidade da implementação	34
3.4 RESULTADOS E DISCUSSÕES.....	36
4 CONCLUSÕES.....	39
4.1 EXTENSÕES	39
REFERÊNCIAS	41

1 INTRODUÇÃO

Dados do Instituto Brasileiro de Geografia e Estatística (IBGE) mostram que no Brasil existem cerca de 9,7 milhões de pessoas com algum tipo de deficiência auditiva, das quais, cerca de 345 mil não conseguem ouvir de modo algum (SECRETARIA NACIONAL DE PROMOÇÃO DOS DIREITOS DA PESSOA COM DEFICIÊNCIA, 2010). Para essas pessoas, uma eficiente forma de comunicação alternativa à fala são as línguas de sinais.

As línguas de sinais são, portanto, consideradas pela linguística como línguas naturais ou como um sistema linguístico legítimo e não como um problema do surdo ou como uma patologia de linguagem. Stokoe, em 1960, percebeu e comprovou que a língua dos sinais atendia a todos os critérios linguísticos de uma língua genuína, no léxico, na sintaxe e na capacidade de gerar uma quantidade infinita de sentenças. (QUADROS; KARNOPP, 2004, p. 30).

No Brasil, a língua de sinais utilizada é denominada Língua BRAsileira de Sinais (LIBRAS), que, devido a grande complexidade que possui, acaba sendo a língua materna de grande parcela dos surdos. Seu ensino é apoiado no Decreto Federal 5626, de 22 de dezembro de 2005 (BRASIL, 2005), que estabelece a obrigatoriedade, desde a educação infantil, do ensino da LIBRAS e também da Língua Portuguesa, como segunda língua para alunos surdos. Na prática muitos surdos dominam menos, por vezes nada, da Língua Portuguesa. Com isso, a comunicação entre pessoas surdas com pessoas sem deficiência auditiva normalmente é dificultosa, e muitos grupos de pais ainda têm resistência em alfabetizar seus filhos surdos com LIBRAS, o que compromete ainda mais o aprendizado destes indivíduos. Por isso, Santana (2007) afirma a necessidade de a sociedade realizar esforços para reconhecer surdez como uma diferença, não como deficiência, e reafirma a importância social de um indivíduo surdo conhecer LIBRAS.

Conferir à língua de sinais o estatuto de língua não tem apenas repercussões linguísticas e cognitivas, mas também sociais. Se ser anormal é caracterizado pela ausência da língua e de tudo que ela representa (comunicação, pensamento, aprendizagem, etc.), a partir do momento em que se tem a língua de sinais como língua do surdo, o padrão de normalidade também muda. Ou seja, a língua de sinais legitima o surdo como “sujeito de linguagem” e é capaz de transformar a “anormalidade” em diferença. Isso é resultado de uma luta pela redefinição do que é considerado normal. A ideia de que a surdez é uma diferença traz com ela uma delimitação de esferas sociais: a identidade surda, a cultura surda, a comunidade surda. (SANTANA, 2007, p. 33).

Diante do exposto, foi construído um protótipo de ferramenta capaz de traduzir textos escritos em Língua Portuguesa para LIBRAS. Para tanto, foi definido um processo para converter sentenças escritas em Língua Portuguesa para sentenças correspondentes em LIBRAS. Após esta conversão, são exibidos os sinais de LIBRAS equivalentes, tomando como base o Dicionário da Língua Brasileira de Sinais (LIRA; SOUZA, 2005).

1.1 OBJETIVOS

O objetivo deste trabalho é desenvolver um protótipo de uma ferramenta para traduzir textos escritos em Língua Portuguesa para LIBRAS.

Os objetivos específicos do trabalho são:

- a) processar textos escritos de acordo com as normas gramaticais da Língua Portuguesa;
- b) traduzir as sentenças de entrada em sinais de LIBRAS, usando o Dicionário da Língua Brasileira de Sinais.

1.2 ESTRUTURA

Este trabalho está dividido nos seguintes capítulos: introdução, fundamentação teórica, desenvolvimento e conclusão. No próximo capítulo são abordados conceitos utilizados para a construção do trabalho, tais como a gramática de LIBRAS, o Dicionário da Língua Brasileira de Sinais, processamento de linguagem natural, relacionando também alguns trabalhos correlatos. O terceiro capítulo trata do desenvolvimento do protótipo, mostrando requisitos, diagramas e especificação, além de trechos da implementação, bem como resultados obtidos. No capítulo final são expostas conclusões e sugestões para extensões futuras.

2 FUNDAMENTAÇÃO TEÓRICA

Os conteúdos que se encontram nas próximas seções foram divididos conforme estudados para o desenvolvimento do protótipo. A seção 2.1 descreve conceitos relacionados a LIBRAS e sua gramática. Na seção 2.2 é abordado o Dicionário da Língua Brasileira de Sinais. A seção 2.3 trata do processamento de linguagem natural. Na seção 2.4 são apresentados trabalhos correlatos.

2.1 LIBRAS

LIBRAS é uma linguagem natural complexa, composta de suas próprias regras gramaticais, que incluem aspectos fonológicos, morfológicos, sintáticos e semânticos, esses, por vezes, distintos dos da Língua Portuguesa. Além disso, LIBRAS é uma língua visual espacial, o que significa que seu uso é realizado através de gestos e expressões com as mãos, face e corpo.

Para traduzir sentenças da Língua Portuguesa para equivalentes em LIBRAS, deve-se usar o sistema de transcrição utilizado pela Federação Nacional de Educação e Integração dos Surdos (FENEIS), que é a forma de transcrição mais difundida para representar LIBRAS em forma textual.

2.1.1 Sistema de transcrição

Lourenço e Meireles (2012) explicam as seguintes convenções:

- a) palavras da Língua Portuguesa com um sinal equivalente em LIBRAS são escritas em letras maiúsculas, como por exemplo, PALAVRA, ANDAR;
- b) duas ou mais palavras da Língua Portuguesa representadas por um único sinal são escritas separadas por hífen, como em SABER-NÃO, CORTAR-CABELO;
- c) palavras da Língua Portuguesa representadas por mais de um sinal têm as palavras que representam cada sinal separadas pelo símbolo ^. Cita-se como exemplo a palavra MÃE, representada pelos seguintes símbolos MULHER^BENÇÃO;
- d) palavras que não possuem sinais equivalentes são representadas escrevendo cada letra separada por um hífen. Assim, tem-se M-A-U-R-O, I-T-A-L-I-A;
- e) é possível soletrar sinais, quando se quer fazer empréstimos da Língua Portuguesa. A transcrição é escrita como no item anterior, porém com as letras em itálico, estas sendo equivalentes ao timbre das palavras, o que acaba por muitas vezes suprir letras, sendo *N-U-N-C-A* a representação para nunca, e *A-H-O* para alho;
- f) não há distinção de gênero e número em LIBRAS de tal forma que na

representação das palavras, a letra que distingue o gênero é substituída pelo símbolo @, como EL@ para representar ele, ela, eles e elas.

2.1.2 Classes gramaticais

Segundo Januário, Leite e Koga (2010, p. 11), “as classes ou categorias gramaticais correspondem a paradigmas sobre um corpo de palavras. Através desses paradigmas se obtêm novos lexemas a partir de outros”. As classes gramaticais de LIBRAS obedecem aos seguintes paradigmas (FELIPE, 1997 apud JANUÁRIO; LEITE; KOGA, 2010):

- a) verbos: existem dois tipos de verbos, os que realizam concordância com algum elemento e os que não realizam. Quanto ao último grupo, estes estão sempre no infinitivo. Já os que têm concordância, são divididos em três categorias:
 - concordância número-pessoal: quando a orientação do discurso muda de acordo com os sujeitos da oração. Nestes casos o verbo é escrito com numerações referentes aos sujeitos. Por exemplo, na frase “Eu pergunto a você.”, o verbo perguntar é representado como 1sPERGUNTAR2s, já na frase “Você me pergunta.”, a representação é 2sPERGUNTAR1s,
 - concordância de gênero: grupo de verbos onde a configuração da mão se altera de acordo com o gênero, que pode ser referente a pessoa, animal ou objeto. O verbo andar está neste grupo, possuindo sinais específicos para se referir a pessoa, animal ou objeto,
 - concordância com a localização: quando a ação do verbo começa ou termina em determinado lugar, como por exemplo, quando se quer indicar que algum objeto foi colocado sobre uma mesa, deve-se fazer o sinal da mesa em determinado ponto e na sequência o do verbo COLOCAR, porém tendo como destino o local onde o sinal da mesa foi realizado, justamente para criar a ideia de localização no espaço;
- b) classificadores: representados por configurações de mão relacionadas a pessoas, animais e objetos, para descrever suas formas. São parte também dos verbos com concordância de gênero e têm como função possibilitar que a construção de sentenças em LIBRAS seja realizada abrangendo maiores detalhes. Por exemplo, se o verbo andar for utilizado para descrever que duas pessoas andaram, é possível realizar o sinal de andar com as duas mãos simultaneamente, ou repetir o símbolo duas vezes. Essa simultaneidade e repetição são casos de utilização prática de um classificador;

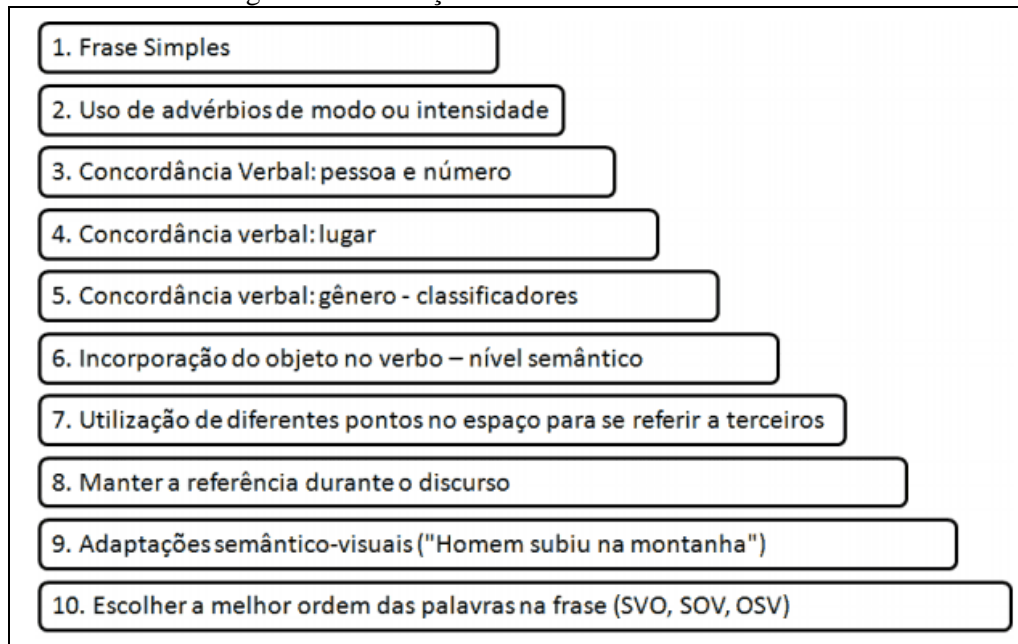
- c) artigos: ao contrário da Língua Portuguesa, em LIBRAS os artigos não são representados;
- d) pronomes: dividem-se em:
 - pronomes pessoais: em LIBRAS é possível contabilizar a quantidade de pessoas num discurso, contendo as classes singular, dual, trial, quatrial e plural. Na prática, a representação dos pronomes da primeira pessoa fica assim: EU, NÓS-2, NÓS-3, NÓS-4, NÓS-GRUPO E NÓS-TOD@,
 - pronomes demonstrativos e advérbios de lugar: os dois grupos possuem os mesmos sinais. Por exemplo, EST@ e AQUI são representados com o mesmo símbolo e, além do uso de tal símbolo, também se pode apontar para o lugar apropriado acompanhado de um olhar para o mesmo,
 - pronomes interrogativos: possuem regras específicas, de acordo com cada símbolo. Os sinais QUE, QUEM, POR-QUE geralmente são realizados no começo da frase, enquanto que QUAL, COMO, PARA-QUE, ONDE e QUEM são utilizados no final da frase;
- e) adjetivos: geralmente são representados após os substantivos que eles caracterizam;
- f) advérbios: não existem símbolos para advérbios de intensidade e modo. Para que seus significados sejam inseridos no contexto de LIBRAS, os sinais aos quais estes tipos de advérbios podem se referir tem suas representações modificadas em relação à velocidade ou sendo repetido algumas vezes. Já os advérbios relacionados a tempo (ontem, hoje, amanhã, entre outros) normalmente são inseridos no início da frase.

2.1.3 Sintaxe

Em LIBRAS, a ordem base presente nas sentenças é a Sujeito-Verbo-Objeto (SVO). Porém, existem outras estruturas aceitas, mesmo que com restrições, que são a OSV e SOV. Tais restrições existem devido à exigência de concordância entre verbo e sujeito/objeto, além da presença de marcas não manuais (JANUÁRIO; LEITE; KOGA, 2010).

Januário, Leite e Koga (2010) desenvolveram uma escala de complexidade em relação à estrutura de frases (em nível sintático) da gramática de LIBRAS. A Figura 1 mostra esta escala.

Figura 1 - Graduação de LIBRAS - nível de frases



Fonte: Januário, Leite e Koga (2010, p. 14).

Conforme as frases contêm elementos característicos de LIBRAS, tais como verbos que necessitam de concordância com lugar ou que precisam que algum sinal referencie um local específico, mais complexa será a transcrição entre as línguas, sem que a tradução deixe de abranger em todos os aspectos o conteúdo da sentença original. Convém destacar que, simultaneamente com um sinal, são realizadas diferentes expressões faciais e corporais, de acordo com o tipo de frase, tendo as seguintes expressões (BRITO, 2014):

- a) forma afirmativa: a expressão facial é neutra;
- b) forma interrogativa: sobrancelhas franzidas e um ligeiro movimento da cabeça inclinando-se para cima;
- c) forma exclamativa: sobrancelhas levantadas e um ligeiro movimento da cabeça inclinando-se para cima e para baixo;
- d) forma negativa: pode ser feita com o acréscimo do sinal NÃO à frase afirmativa, com a incorporação de um movimento contrário ao do sinal negado, com um aceno na cabeça que pode ser feito simultaneamente com a ação que está sendo negada.

2.2 DICIONÁRIO DA LÍNGUA BRASILEIRA DE SINAIS

O Dicionário da Língua Brasileira de Sinais (LIRA; SOUZA, 2005) teve sua concepção em agosto de 1997, no Instituto Nacional de Educação de Surdos (INES), com o intuito de desenvolver um vocabulário Língua Portuguesa versus LIBRAS acessível via internet.

Uma versão esteve disponível ao público em caráter experimental no ano de 1999, contendo cerca de mil vocábulos, todos com seus respectivos vídeos para representação do sinal equivalente em LIBRAS. Este foi um grande diferencial em relação aos materiais existentes, pois assim o sinal era visto a partir de todos os parâmetros que o compõe: configuração de mãos, orientação / direcionalidade, ponto de articulação e movimento. Foi uma experiência pioneira no Brasil que, devido à repercussão que obteve e empenho dos envolvidos, ganhou recursos para ser aperfeiçoado e de fato ser formalizado como um dicionário de LIBRAS. Em março de 2001 o dicionário teve sua primeira versão concluída.

Em fevereiro de 2005 é finalizada a segunda versão do dicionário, até agora a mais recente, contando com mais de 5800 sinais, além de melhoria na qualidade dos vídeos que os representam. Esta mesma versão está disponível de forma *online* desde 2008 (Figura 2).

Figura 2 - Versão *online* do Dicionário da Língua Brasileira de Sinais

Fonte: Lira e Souza (2008).

Entre suas funcionalidades, está a possibilidade de pesquisar sinais por ordem alfabética, por assunto e também pela configuração da mão, deixando as duas línguas em estado de igualdade. Para cada sinal (vídeo), há a definição da palavra (acepção), bem como a que assunto ela pertence, qual a configuração da mão e a classe gramatical, além de uma frase exemplo em português em que a palavra é utilizada, trazendo a sua transcrição em LIBRAS. Importante ressaltar que nos casos que uma palavra possui mais de um sinal, como no caso de “direito”, que tem sinais referentes para quando significa curso, lado e certo, o dicionário traz

a palavra seguida de uma numeração (“direito1”, “direito2” e assim por diante). Outros motivos para terem palavras com mais de um sinal incluem a palavra ter mais de uma representação em âmbito nacional. Nesses casos, o sinal que vem primeiro é o mais utilizado, além de poder possuir variações regionais, sendo incluídas, devido a restrições de prazo para o desenvolvimento da pesquisa, apenas sinais nacionais e variações do Rio de Janeiro.

2.3 PROCESSAMENTO DE LINGUAGEM NATURAL

A área que estuda os problemas da geração e compreensão automática de linguagens naturais é denominada Processamento de Linguagens Naturais (PLN). Em PLN o material de entrada do processamento é um texto que deve ser analisado, ou seja, recortado em unidades menores para a compreensão completa dos mecanismos de operação envolvidos em cada uma dessas unidades. Assim, o PLN recorre àqueles campos específicos da Linguística, procurando depreender da sua descrição as informações que irão fazer da máquina um instrumento sensível aos fenômenos da língua natural (SILVA et al., 2007, p. 15).

O primeiro passo para realizar o processo de tradução de sentenças da Língua Portuguesa para LIBRAS é a separação das palavras, os chamados *tokens*, para através de uma análise léxica, adaptar os *tokens* para a gramática de LIBRAS. Pode-se citar os seguintes exemplos de adaptações: desconsiderar os artigos, transcrever os verbos no infinitivo, entre outras. Com os léxicos definidos, é realizada a análise sintática, que deverá definir a ordem correta dos *tokens*, de acordo com a gramática de LIBRAS.

Nota-se que os *tokens* são classificados de acordo com a classe gramatical da palavra. Por exemplo, a palavra “computador” é categorizada como substantivo, enquanto a palavra “estudo” poderá ser classificada tanto como substantivo, quanto como verbo, dependendo do contexto em que se encontra nas sentenças. Esse processo de classificação é denominado etiquetagem, que consiste na implementação de *parsers* para identificar nas frases as classes das palavras que as compõem (MÜLLER, 2003, p. 4). No Quadro 1 tem-se um exemplo de uma frase com as palavras etiquetadas.

Quadro 1 - Etiquetagem da frase “Conversei com arquiteto para fazer uma casa moderna.”

token	classe gramatical
Conversei	verbo
com	preposição
arquiteto	substantivo
para	preposição
fazer	verbo no infinitivo
uma	artigo
casa	substantivo
moderna	adjetivo

2.4 TRABALHOS CORRELATOS

A seguir estão relacionados três trabalhos semelhantes à pesquisa apresentada neste trabalho, que são: ProDeaf (PRODEAF TECNOLOGIAS ASSISTIVAS LTDA, 2014), conjunto de softwares com diversas funções para auxílio da tradução de português para LIBRAS; Poli-Libras (JANUÁRIO; LEITE; KOGA, 2010), que trabalha com a conversão de textos em português para animação gráfica em 3D referente aos sinais de LIBRAS; e o LIBROL, software tradutor de português para LIBRAS (CARVALHO et al., 2013), que converte textos em português para a estrutura gramatical de LIBRAS.

2.4.1 PRODEAF

O ProDeaf é uma solução mercadológica constituída por diversos softwares, que promovem acessibilidade para surdos traduzindo textos e voz do português para LIBRAS. Conta com aplicações para diversos setores, entre eles para web e dispositivos móveis, como mostrado na Figura 3, onde é possível utilizar voz ou texto digitado como entrada, produzindo como saída uma animação gráfica em 3D que representa a tradução automática da entrada para LIBRAS.

Figura 3 - ProDeaf para dispositivos móveis



Fonte: ProDeaf Tecnologias Assistivas Ltda (2014).

Outra solução possibilita a tradução automática de *websites* para seu equivalente em LIBRAS, sendo que a empresa também oferece que determinado *site* seja traduzido manualmente, dependendo das necessidades do cliente. Porém, o artefato que mais se aproxima da proposta deste trabalho é o tradutor *online*, pois, além de traduzir pequenos textos de até 140 caracteres para LIBRAS, seguindo de forma mais fiel a gramática de

LIBRAS, o que não ocorre na versão para dispositivos móveis, também permite que os usuários adicionem sinais, configurando novos movimentos para o *avatar* de saída.

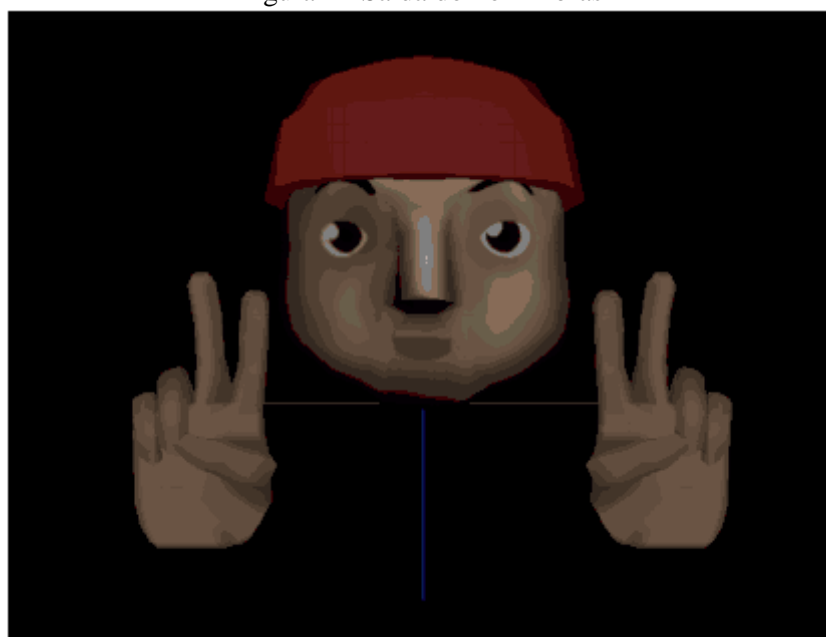
A versão para dispositivos móveis difere consideravelmente da versão *online*, pois, ao contrário da versão *online*, os sinais da versão móvel são limitados, tendo como vantagem que é possível utilizar a ferramenta sem necessitar de conexão com a Internet. O processo de tradução é simplificado na versão móvel, se comparado com a *online*, com muitas das regras da gramática de LIBRAS ignoradas, o que, aliado a quantia de sinais fixos, torna a ferramenta móvel limitada, com muitas frases apenas soletradas.

2.4.2 POLI-LIBRAS

O projeto surgiu como um trabalho de conclusão de curso do Departamento de Engenharia de Computação da Escola Politécnica da Universidade de São Paulo (USP). Foi desenvolvido ao longo do ano de 2010, com o objetivo de realizar uma tradução de frases que leve em conta as diferenças de estruturas sintáticas entre a Língua Portuguesa e LIBRAS, de forma a gerar uma saída mais compreensível aos surdos, representada por animações gráficas em 3D.

O projeto foi desenvolvido em Java, voltado para *web*, com a característica de ser construído em módulos, sendo que todo o seu código fonte é de uso aberto. O módulo que trata da animação gráfica é denominado Virtual Jonah e foi desenvolvido com o auxílio da *Application Programming Interface* (API) Processing. A Figura 4 mostra um exemplo de saída da aplicação.

Figura 4 - Saída do Poli-Libras



Fonte: Januário, Leite e Koga (2010, p. 63).

Os sinais são formalizados através de um modelo de arquivo *Extensible Markup Language* (XML), cuja construção é realizada em um módulo separado das outras partes da aplicação, tornando possível que sejam reutilizados em outros projetos, além de que os usuários finais podem incluir novos sinais, configurando determinados parâmetros. Inicialmente, os desenvolvedores do trabalho pretendiam deixar os arquivos de sinais disponíveis via *webservice*, porém, atualmente tal *webservice* não está mais disponível.

Ao todo, foram desenvolvidas soluções que englobam as seguintes funções, para que a tradução ficasse o mais fiel possível às regras da gramática de LIBRAS: (i) analisador morfológico; (ii) analisador sintático; (iii) dicionário de português para LIBRAS; (iv) sintetizador de sinais; (v) ferramenta colaborativa para criação e edição de sinais.

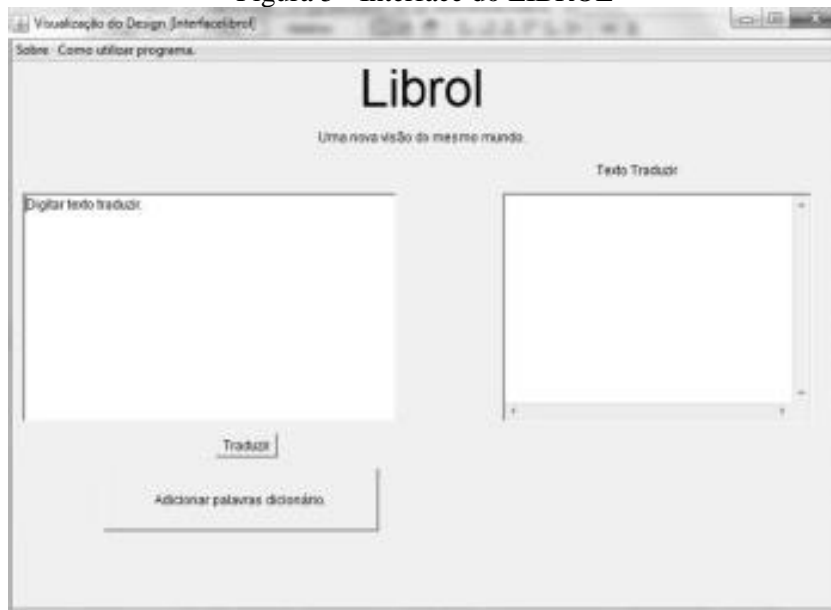
2.4.3 LIBROL

Criado com o intuito principal de facilitar a vida escolar de portadores de deficiência auditiva, segundo Carvalho et al. (2013), o software basicamente é formado por um campo de entrada de texto onde os usuários devem digitar textos em Língua Portuguesa, para este ser convertido para a gramática de LIBRAS. Objetiva melhorar o desempenho de leitura de pessoas que dominam LIBRAS, visto que poderá reduzir o tempo de leitura pela metade, já que, para quem tem LIBRAS com língua materna, se torna mais fácil compreender textos que estão próximos de sua gramática.

O LIBROL inicialmente realiza uma análise morfológica do texto de entrada informado pelo usuário, retirando palavras ou expressões, que serão classificadas de acordo com a classe gramatical que pertencem. Quando a ferramenta identificar verbos, estes terão seus radicais retirados, por meio de expressão regular, que em seguida são comparados e substituídos por seus equivalentes em LIBRAS, que estão armazenadas previamente em um banco de dados.

A ferramenta foi desenvolvida em Java, possuindo também a funcionalidade de adicionar ao seu banco de dados inicial palavras que o usuário identifique como sem utilidade no texto. Desta forma, da próxima vez que este for iniciado, as personalizações realizadas pelo usuário serão reconhecidas e utilizadas numa próxima correspondência de textos. A Figura 5 mostra a interface gráfica do protótipo.

Figura 5 - Interface do LIBROL



Fonte: Carvalho et al. (2013).

3 DESENVOLVIMENTO DO PROTÓTIPO

Este capítulo está dividido em seções que mostram:

- a) requisitos principais do trabalho;
- b) especificação do protótipo;
- c) técnicas, ferramentas utilizadas e operacionalidade da implementação;
- d) resultados obtidos.

3.1 REQUISITOS PRINCIPAIS DO PROBLEMA A SER TRABALHADO

O protótipo deverá:

- a) permitir que o usuário digite o texto que será traduzido (Requisito Funcional - RF);
- b) permitir que o usuário carregue o texto que será traduzido através de um arquivo de texto (RF);
- c) permitir que o usuário salve um texto previamente digitado ou carregado de um arquivo de texto (RF);
- d) transcrever textos escritos em Língua Portuguesa para LIBRAS de acordo com o sistema de transcrição utilizado pela FENEIS (RF);
- e) utilizar o Dicionário da Língua Brasileira de Sinais como fonte dos sinais que serão exibidos (Requisito Não-Funcional - RNF);
- f) ser desenvolvido utilizando a linguagem de programação Java (RNF);
- g) utilizar a biblioteca auxiliar Cogroo para realizar a etiquetagem de palavras (RNF);
- h) utilizar a biblioteca auxiliar Xuggle para realizar a leitura dos sinais que serão exibidos (RNF).

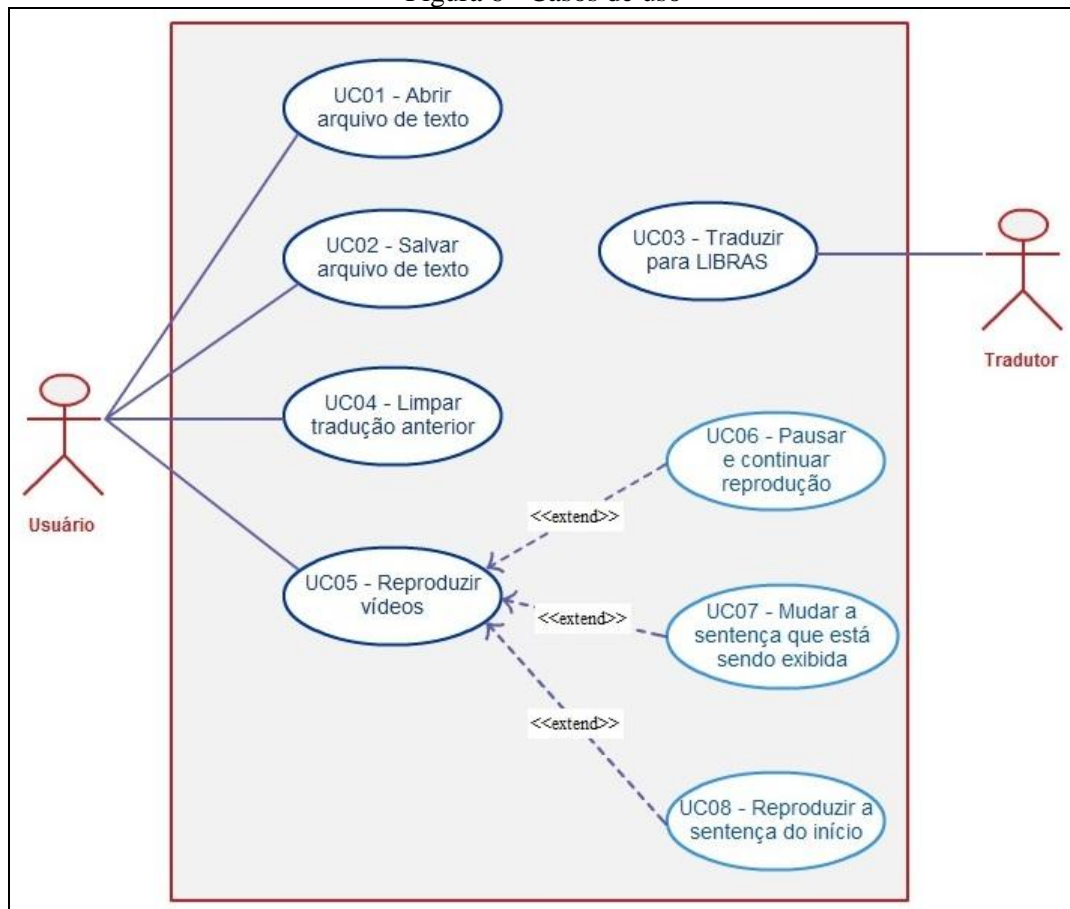
3.2 ESPECIFICAÇÃO

Nessa seção é mostrada a especificação do protótipo, através dos diagramas de casos de uso, de classes e sequência, da *Unified Modeling Language* (UML). Todos os diagramas foram criados com a ferramenta Creately.

3.2.1 Diagrama de casos de uso

O diagrama de casos de uso é apresentado na Figura 6, tendo como atores o usuário final e o tradutor. Ao iniciar o protótipo, o usuário poderá realizar a ação descrita no caso de uso `Abrir arquivo de texto`, detalhado no Quadro 2, ou digitar um texto no campo de entrada de texto.

Figura 6 - Casos de uso



Quadro 2 - Caso de uso: Abrir arquivo de texto

UC01 – Abrir arquivo de texto	
Descrição	Permite selecionar um arquivo de texto.
Atores	Usuário
Cenário principal	<ol style="list-style-type: none"> 1. Usuário pressiona o botão abrir. 2. Usuário escolhe arquivo que deseja carregar. 3. Protótipo carrega o arquivo selecionado e, se existir, carrega também a tradução para LIBRAS do arquivo selecionado.
Pré-condições	O arquivo deve ser no formato texto (<code>txt</code>) e escrito em português.
Pós-condições	Campo de entrada de texto carregado com o texto do arquivo. Se existir, a tradução para LIBRAS do arquivo selecionado deve ser carregada para a memória do protótipo.

Tendo o campo de entrada de texto preenchido, o usuário poderá realizar a requisição do processo de tradução, cuja ação está descrita no caso de uso Traduzir para LIBRAS, detalhado no Quadro 3. Também é possível salvar o texto de entrada (caso de uso Salvar arquivo de texto, detalhado no Quadro 4). Caso já tenha sido realizada a tradução e o texto de entrada não tenha sido modificado desde então, o protótipo cria um arquivo auxiliar que guardará as informações da tradução. Esse arquivo será lido posteriormente quando o arquivo for novamente aberto.

Quadro 3 - Caso de uso: Traduzir para LIBRAS

UC03 – Traduzir para LIBRAS	
Descrição	Permite traduzir textos em português para LIBRAS.
Atores	Tradutor
Cenário principal	1. Usuário pressiona o botão traduzir. 2. Tradutor traduz o texto de entrada para sentenças em LIBRAS.
Pré-condições	O campo de entrada de texto não pode estar vazio e deve ser escrito em português.
Pós-condições	Sentença(s) traduzida(s) para LIBRAS, podendo ser reproduzida(s).

Quadro 4 - Caso de uso: Salvar arquivo de texto

UC02 – Salvar arquivo de texto	
Descrição	Permite salvar o texto do campo de entrada em um arquivo externo.
Atores	Usuário
Cenário principal	1. Usuário pressiona o botão salvar. 2. Usuário informa nome e local em que deseja salvar o arquivo. 3. Protótipo salva o texto de entrada com o nome e no local informado e, se existir uma tradução do texto do campo de entrada e se o texto não foi modificado depois da execução da tradução, salva arquivo externo auxiliar contendo a tradução do texto de entrada.
Pré-condições	O campo de entrada de texto não pode estar vazio e deve ser escrito em português.
Pós-condições	Texto de entrada e tradução salvos.

O Quadro 5 detalha o caso *Limpar tradução anterior*, que limpará o campo de entrada de texto, caso esteja preenchido, e limpará, caso tenha, a tradução para LIBRAS da memória do aplicativo.

Quadro 5 - Caso de uso: Limpar tradução anterior

UC04 – Limpar tradução anterior	
Descrição	Permite limpar o campo de entrada de texto e a tradução gerada.
Atores	Usuário
Cenário principal	1. Usuário pressiona o botão limpar. 2. Protótipo apaga o texto de entrada e remove tudo relacionado à tradução.
Pré-condições	O campo de entrada de texto deve estar preenchido ou o processo de tradução deve ter sido executado ao menos uma vez.
Pós-condições	Campo de entrada de texto e memória do protótipo limpos.

Por fim, tendo efetuado a tradução, o usuário terá a oportunidade de realizar a ação *Reproduzir vídeos*, mostrado em detalhes no Quadro 6, que poderá ser acompanhado das ações *Pausar e continuar reprodução*, *Mudar a sentença que está sendo exibida*, *Reproduzir a sentença do início*, detalhados nos Quadro 7, Quadro 8 e Quadro 9, respectivamente.

Quadro 6 - Caso de uso: Reproduzir vídeos

UC05 – Reproduzir vídeos	
Descrição	Permite reproduzir vídeos com os sinais de LIBRAS.
Atores	Usuário
Cenário principal	1. Usuário pressiona o botão reproduzir. 2. Protótipo inicia a exibição das sentenças em LIBRAS. 3. Usuário assiste a reprodução dos vídeos.
Pré-condições	O processo de tradução deve ter sido executado ao menos uma vez.
Pós-condições	Vídeos são exibidos com as sentenças em LIBRAS.

Quadro 7 - Caso de uso: Pausar e continuar reprodução

UC06 – Pausar e continuar reprodução	
Descrição	Permite pausar a reprodução dos vídeos com os sinais de LIBRAS.
Atores	Usuário
Cenário principal	1. Usuário pressiona o botão pausar. 2. Protótipo pausa a reprodução. Caso o botão pausar seja pressionado novamente, a reprodução continua.
Pré-condições	A reprodução dos vídeos deve estar em execução.
Pós-condições	Reprodução é pausada / continuada.

Quadro 8 - Caso de uso: Mudar a sentença que está sendo exibida

UC07 – Mudar a sentença que está sendo exibida	
Descrição	Permite trocar de sentença sem a necessidade de reproduzi-la por completo.
Atores	Usuário
Cenário principal	1. Usuário pressiona o botão avançar sentença ou o botão voltar sentença. 2. Protótipo muda a exibição de sentença.
Pré-condições	O processo de tradução deve ter sido executado ao menos uma vez, com mais de uma sentença gerada.
Pós-condições	Sentença exibida é alterada.

Quadro 9 - Caso de uso: Reproduzir a sentença do início

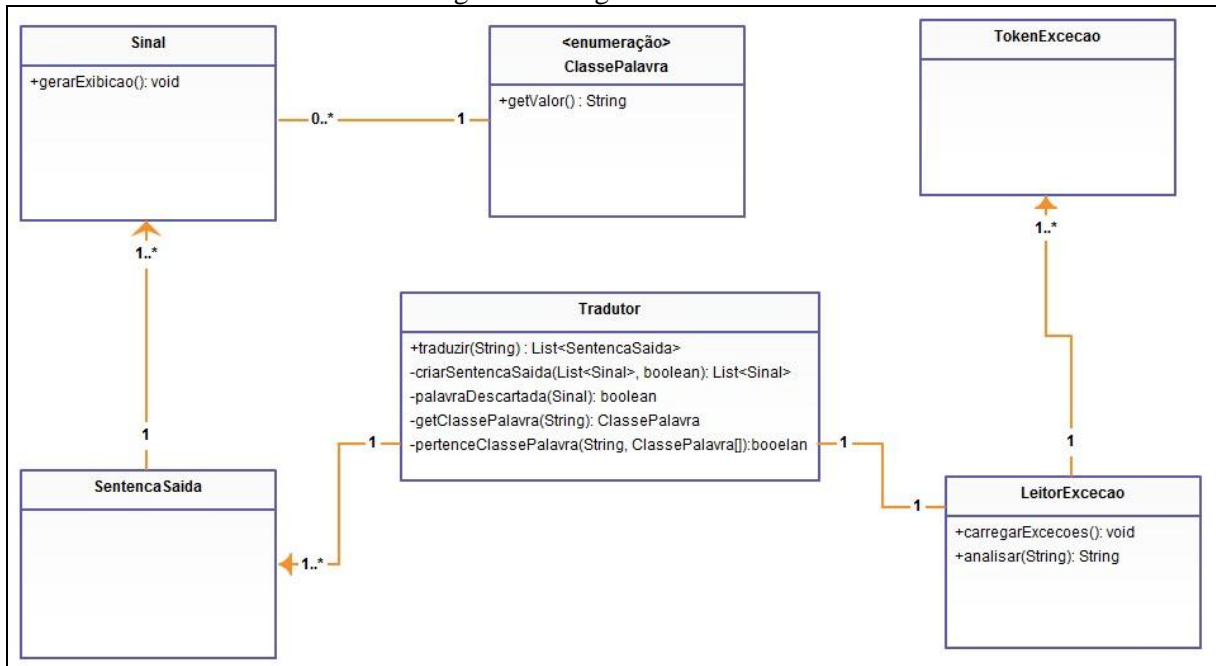
UC08 – Reproduzir sentença do início	
Descrição	Permite reinicia a reprodução da sentença em LIBRAS.
Atores	Usuário
Cenário principal	1. Usuário pressiona o botão reiniciar. 2. Protótipo reinicia a reprodução da sentença.
Pré-condições	A reprodução dos vídeos deve estar em execução.
Pós-condições	Sentença volta a ser reproduzida do começo.

3.2.2 Diagrama de classes

A Figura 7 apresenta o diagrama de classes do protótipo. Assim que o protótipo é iniciado, o método `carregarExcecoes` da classe `LeitorExcecao` é executado, carregando as regras do dicionário de exceções, para a classe `TokenExcecao`. Quando o processo de tradução for iniciado, o método `traduzir` da classe `Tradutor` será executado. Inicialmente, este método chamará o método `analisar` da classe `LeitorExcecao`, para verificar se o texto informado pelo usuário possui algum trecho em que alguma regra inserida no dicionário de exceções se aplica. Caso tenha, o método retornará o texto de entrada com as devidas substituições. Após isso, o método `traduzir` realizará a etiquetagem de palavras usando métodos e classes da biblioteca auxiliar `Cogroo`, cujos atributos e objetos relevantes serão

convertidos para as classes `Sinal` e `SentencaSaida`, as quais contam também com outros atributos, sendo um para verificar se o sinal já foi analisado na tradução e outro para guardar o caminho do vídeo que deverá ser exibido, ou caso não exista, o caminho das letras que serão soletradas. Esse processo é realizado no método `gerarExibicao` da classe `Sinal`. Então, o método `traduzir` devolverá para a interface gráfica do protótipo uma lista de objetos da classe `SentencaSaida`, contendo a tradução para LIBRAS do texto solicitado pelo usuário.

Figura 7 - Diagrama de classes



Durante a execução do método `traduzir`, os seguintes métodos privados da classe `Tradutor` são executados:

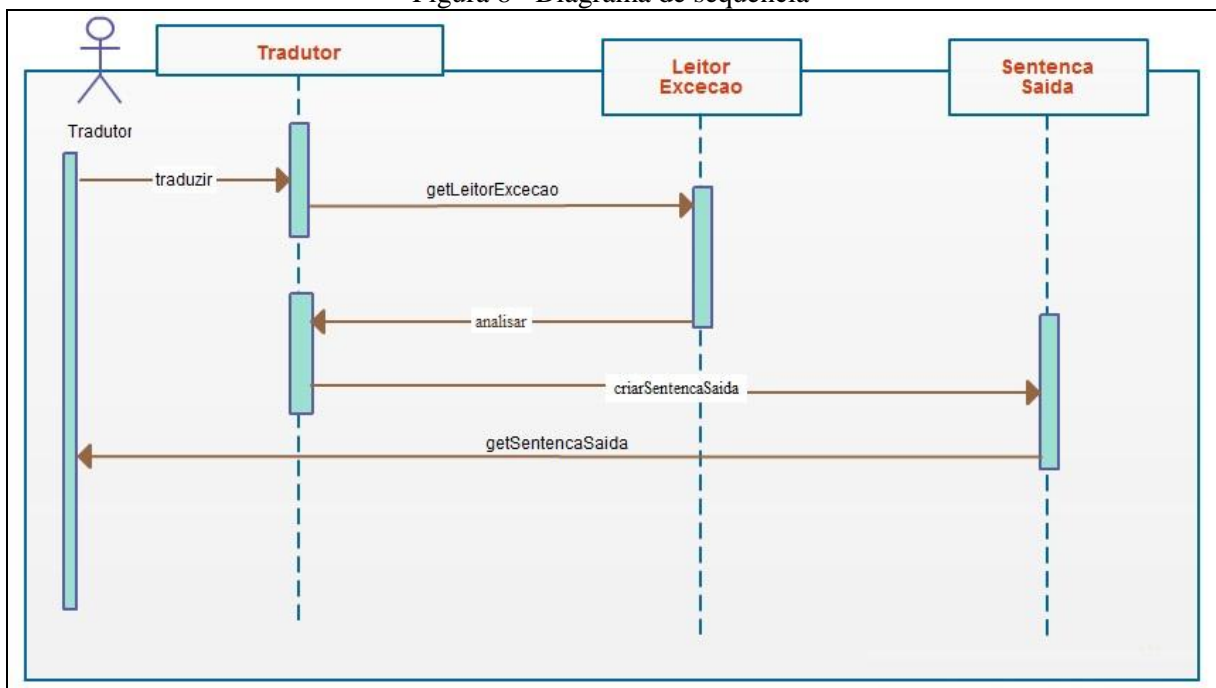
- `palavraDescartada`: que descarta palavras que pertencem a classes gramaticais que são usadas em LIBRAS, mas que dependendo do contexto, não devem ser exibidas. Exemplos incluem os verbos `ser` e `estar`, como observado na frase “Eu estou cansado.”, cuja transcrição para LIBRAS é “EU CANSAD@”. Nesse método também é verificado se existe algum verbo no particípio ou gerúndio, pois nesses casos existe a possibilidade de se descartar o verbo anterior a este, como na frase “Ele fica pensando na vida.”, onde o verbo “ficar” não deverá ser traduzido para LIBRAS, com a transcrição correspondente sendo “EL@ PENSAR VIDA”;
- `getClassePalavra`: que retorna a classe gramatical de um `token`;
- `pertenceClassePalavra`: que verifica se a classe gramatical de um sinal pertence a uma lista de classes gramaticais, retornando verdadeiro em caso positivo;
- `criarSentencaSaida`: que faz os ajustes finais na tradução antes de retornar

objetos da classe `SentencaSaida`. Entre os ajustes realizados, cita-se: (i) verificar se existe sujeito na sentença e, caso seja oculto, inserir um pronome pessoal relativo à pessoa na transcrição para LIBRAS; (ii) caso a sentença possua uma negação, inverter a ordem dos sinais; (iii) caso seja uma sentença interrogativa e essa não possua as palavras “onde”, “por que”, “por quê”, “qual”, “quando” e “quem”, adicionar o sinal “perguntar” no início; (iv) realiza o tratamento de numerais, tanto para números escritos por extenso como para números escritos em algarismos, inserindo os sinais dos números e relacionados para indicar dezena, centena, milhar, milhão e bilhão.

3.2.3 Diagrama de sequência

A Figura 8 mostra o diagrama de sequência, detalhando a ordem das ações do tradutor. No início, será executado o método `traduzir`, que receberá do `LeitorExcecao` o texto de entrada analisado e, caso necessário, alterado, de acordo com as regras do dicionário de exceções. Em seguida, serão criadas as sentenças de saída, contendo as sentenças em LIBRAS, que são retornadas para o tradutor.

Figura 8 - Diagrama de sequência



3.3 IMPLEMENTAÇÃO

A seguir são mostradas as técnicas e ferramentas utilizadas, bem como a operacionalidade do protótipo.

3.3.1 Técnicas e ferramentas utilizadas

O protótipo foi desenvolvido em Java, com a versão 6, tendo a IDE Eclipse, na versão Luna (4.4), como ambiente de desenvolvimento, a biblioteca Swing para criação da interface gráfica e as bibliotecas auxiliares Cogroo na versão 4.0 e Xuggle na versão 5.4.

Cogroo é usada para realizar a etiquetagem das palavras. A biblioteca processa textos retornando as sentenças com seus respectivos *tokens* e a classe gramatical que cada um pertence. O Quadro 10 mostra um trecho de código exemplificando a utilização da biblioteca e, em seguida, a Figura 9 mostra a saída da execução desse trecho de código.

Quadro 10 - Trecho de código mostrando a utilização de Cogroo

```
// percorre lista de sentenças
for (Sentence sentence : document.getSentences()) {
    // imprime texto da sentença
    System.out.println(sentence.getText());

    for (Token token : sentence.getTokens()) { // lista de tokens
        try { // um array com os possíveis lemas para o par lexeme+postag
            System.out.println(token.getLemmas()[0]);
        } catch (Exception e) {
            // caso não tenha nenhum lema, imprime o lexema
            System.out.println(token.getLexeme());
        }
        // imprime a classe morfológica de acordo com o contexto
        System.out.println(token.getPOSTag());
        // retorna a tag sintática do token
        // (sujeito, predicado, objeto direto, entre outros)
        token.getSyntacticTag();
        // retorna gênero, número, tempo, entre outros
        token.getFeatures();
    }
}
```

Figura 9 - Saída da execução do trecho de código mostrado no Quadro 10

```
Conversei com arquiteto para fazer uma casa moderna.
conversar
v-fin
com
prp
arquiteto
n
para
prp
fazer
v-inf
um
art
casa
n
moderno
adj
.
.
```

Xuggle é utilizado para processamento de vídeos. É através dele que são carregados e reproduzidos os vídeos com os sinais em LIBRAS. Os vídeos se encontram no formato *swf*, que é um *container* para arquivos multimídia.

3.3.2 Tradução de sentenças de Língua Portuguesa para LIBRAS

Durante o desenvolvimento do protótipo, teve-se a necessidade de criar um mecanismo para realização de um pré-processamento de texto, antes da análise da tradução propriamente dita. Esse mecanismo chama-se dicionário de exceções e está organizado em um arquivo texto (formato `txt`), que contém diversas regras. Basicamente, este mecanismo teve que ser criado para que o processo de tradução ocorra de forma correta nas seguintes situações:

- a) em casos de palavras que não possuem um símbolo em LIBRAS, porém o sinônimo possui, então é criada uma regra no dicionário que substitui tal palavra por seu sinônimo, como a palavra “garoto”, cujo símbolo “menino” poderá ser usado para representá-la;
- b) em casos de palavras que a tradução para LIBRAS é um sinal composto, ou seja, sua tradução equivale a dois ou mais sinais, que separados possuem significado diferente, como por exemplo a palavra “mãe”, que em LIBRAS é representada com os sinais “mulher” e “benção”. Nesse caso, é adicionada uma regra no dicionário de exceções para quando tiver a palavra “mãe” substituir por “mulher benção”;
- c) em casos que se observou erros na etiquetagem de palavras realizada pelo Cogroo, sobretudo quando uma palavra pode ser verbo ou substantivo, dependendo do contexto. Por exemplo, a palavra “estudo”, que mesmo em frases que é verbo, como em “eu estudo na FURB”, é classificada como substantivo. Então, para corrigir essa situação, foi criada uma regra no dicionário de exceções para quando um texto contiver o trecho “eu estudo”, substituir por “eu estudar”, visto que em LIBRAS os sinais para verbos sempre estão no infinitivo;
- d) em casos de erros gramaticais, como “guarda-chuva”, por exemplo, que se escrito sem hífen, será traduzido com os sinais de LIBRAS para “guarda” e “chuva”, que é diferente do sinal de “guarda-chuva”. Para corrigir, basta criar uma regra para quando as palavras “guarda” e “chuva” estiverem juntas em algum texto, trocar por “guarda-chuva”.

Devido à complexidade e ao fato de que cada regra deve ser inserida com certo cuidado, decidiu-se não deixar o dicionário de exceções editável via *software* pelo usuário final, devendo ser ampliado diante de alguma necessidade vista por um especialista. A Figura 10 mostra a estrutura interna do arquivo do dicionário de exceções.

Figura 10 - Estrutura do dicionário de exceções

```

estudo * eu $ | eu estudar
guarda * $ chuva | guarda-chuva
mãe * $ | mulher benção
mae * $ | mulher benção

```

Cada linha do arquivo é uma regra. No momento que o usuário iniciar a execução do protótipo, é realizado o *parser* de todas as regras para a memória do protótipo. Para gerenciamento das regras, foi elaborada uma estrutura que posteriormente é interpretada no protótipo. Ela funciona da seguinte maneira: contém uma palavra-chave, que será comparada com todas as palavras do texto de entrada (para fins de otimização, o dicionário deve ser montado com as palavras chave em ordem alfabética, pois antes do protótipo analisar a próxima regra, é verificado se a palavra-chave atual é menor que a da próxima regra); se a palavra-chave for igual, então é analisada se a sentença antes e depois da palavra da regra é igual ao do texto de entrada; se for igual, o trecho do texto de entrada que é igual a sentença da regra é substituída pela nova sentença. Nas regras do dicionário, a separação da palavra-chave, sentença anterior e posterior (que são opcionais), nova sentença, é feita com os caracteres especiais “*”, “\$”, “|”, respectivamente, como mostrado no exemplo do Quadro 11.

Quadro 11 - Exemplo de regras do dicionário de exceções

palavra-chave	*	sentença anterior	\$	sentença posterior		nova sentença
estudo		eu		-		eu estudar
guarda		-		chuva		guarda-chuva
mãe		-		-		mulher benção
mae		-		-		mulher benção

Deve-se ressaltar que não será possível tratar todos os erros de etiquetagem gerados pelo Cogroo. Por exemplo, na frase “Eu ainda estudo na FURB.”, o verbo “estudo” será classificado como substantivo e a condição da primeira regra mostrada no Quadro 11 não será verdadeira nessa sentença.

Após verificar se alguma regra do dicionário de exceções é verdadeira, substituindo os trechos onde as condições são atendidas, o protótipo iniciará o processo de tradução propriamente dito, inicialmente realizando a etiquetagem de palavras, gerando os *tokens* e removendo as palavras que pertencem a classes gramaticais que não são consideradas em LIBRAS. O Quadro 12 apresenta o trecho de código onde é implementado esse processo, descartando artigos, preposições e conjunções coordenativas.

Quadro 12 - Trecho de código onde são descartados *tokens* não utilizados em LIBRAS

```

01 boolean naoReconhecer = false;
02 if (token.getPOSTag().equalsIgnoreCase(ClassePalavra.ARTIGO.getVALOR())) {
03     naoReconhecer = true;
04 }
05 if (token.getPOSTag().equalsIgnoreCase(ClassePalavra.PREPOSICAO.getVALOR())) {
06     naoReconhecer = true;
07 }
08 if (token.getPOSTag().equalsIgnoreCase
09     (ClassePalavra.CONJUNCAO_COORDENATIVA.getVALOR())
10     && token.getLexeme().equalsIgnoreCase("e")) {
11     naoReconhecer = true;
12 }
13 if(token.getSyntacticTag().contains("SUBJ")) {
14     achouSujeito = true;
15 }

```

Após descartar os *tokens* que não têm equivalência em LIBRAS, a tradução prossegue, primeiramente verificando se a sentença possui sujeito (verificado através da *tag* sintática de cada *token* gerado na etiquetagem, mostrado na linha 13 do Quadro 12). Caso a frase em português possua sujeito oculto, é verificado através do verbo qual é a pessoa, para então adicionar o sinal do sujeito na sentença em LIBRAS, como mostrado no Quadro 13. Da mesma forma, se for uma frase interrogativa e não possuir um pronome (qual, quem, onde, entre outros), é adicionado o sinal “perguntar” no início da sentença em LIBRAS.

Em relação ao trecho de código apresentado no Quadro 13, ele é executado apenas se a primeira palavra da sentença for um verbo. As condições principais determinam em que pessoa esse verbo se encontra. Observa-se que, apesar de não existir distinção de gênero e número em LIBRAS, na primeira pessoa do plural é inserido o sinal “NÓS”, que é diferente do sinal “EU”. Nota-se ainda que, caso seja um verbo que esteja conjugado na terceira pessoa do singular (linha 6), é necessário verificar se a frase contém sujeito indeterminado (caracterizado pela partícula “-se” após o verbo, na terceira pessoa do singular), que é diferente de sujeito oculto, e logo, não deverá ser adicionado o sinal. Por fim, na linha 19 é inserido um novo sinal com o sujeito (caso não seja sujeito indeterminado) na lista de sinais da sentença.

Seguindo, caso algum dos *tokens* pertença às classes de verbo gerúndio ou particípio, é realizada uma remoção, caso necessário, de outros verbos. Por exemplo, na frase “Eu estou estudando na FURB.”, o verbo “estar” não deve aparecer na sentença em LIBRAS, pois apenas com “estudar” a sentença já faz sentido. Casos semelhantes ocorrem em outras situações com outras classes de verbo, por exemplo, em “Ele está triste.”, a sentença em LIBRAS não necessita incluir o verbo “estar”, sendo necessário removê-lo da tradução.

Continuando, quando o *token* “não” estiver em alguma frase, seguido de um verbo, na tradução para LIBRAS, os dois trocam de lugar na sentença. Em relação a numerais,

independente se escrito em algarismos ou por extenso, o protótipo traduzirá de acordo com as regras da gramática de LIBRAS.

Quadro 13 - Trecho de código que realiza a inserção do sujeito na sentença

```

01 String pessoa = "";
02 if(sentenca.get(0).getCaracteristica().matches(".*1S.*")) {
03     pessoa = "eu";
04 } else if(sentenca.get(0).getCaracteristica().matches(".*2S.*")) {
05     pessoa = "você";
06 } else if(sentenca.get(0).getCaracteristica().matches(".*3S.*")) {
07     if(!(sentenca.size() > 1
08         && sentenca.get(1).getToken().equalsIgnoreCase("- se"))) {
09         pessoa = "ele";
10     }
11 } else if(sentenca.get(0).getCaracteristica().matches(".*1P.*")) {
12     pessoa = "nós";
13 } else if(sentenca.get(0).getCaracteristica().matches(".*2P.*")) {
14     pessoa = "você";
15 } else if(sentenca.get(0).getCaracteristica().matches(".*3P.*")) {
16     pessoa = "ele";
17 }
18 if(!pessoa.isEmpty()) {
19     sentenca.add(0, new Sinal(pessoa, new String[{}],
20         ClassePalavra.PRONOME_PESSOAL 20, ""));
21 }

```

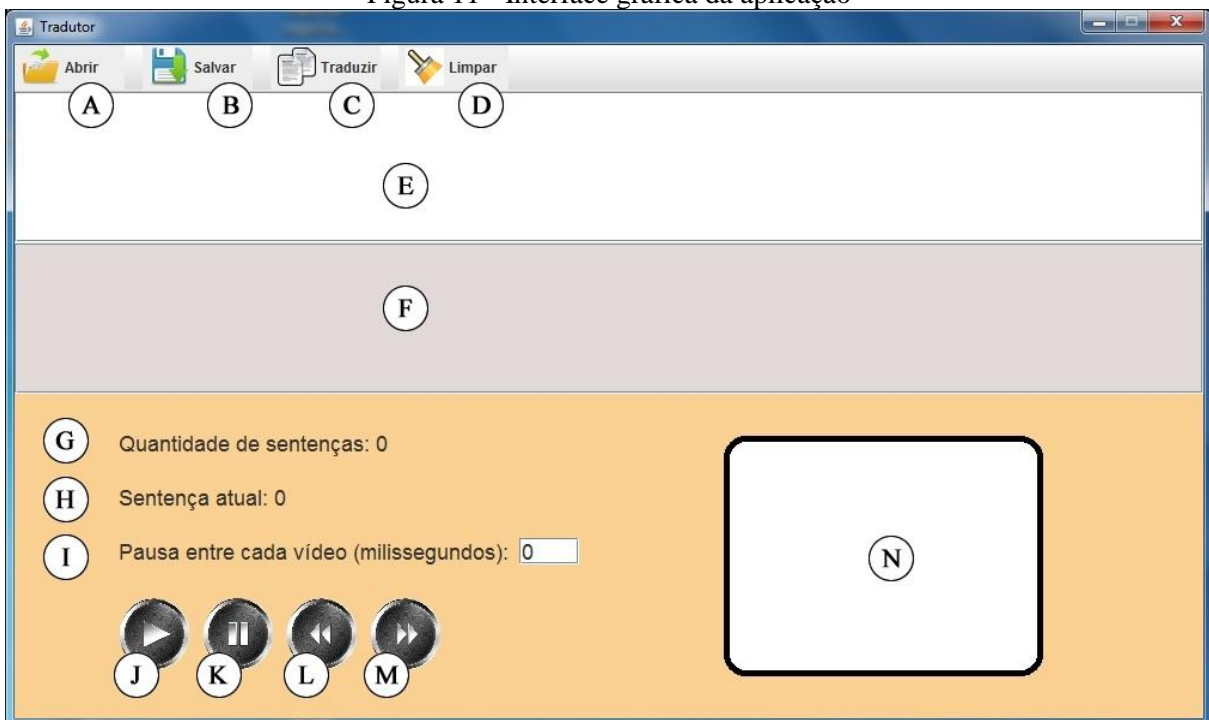
3.3.3 Operacionalidade da implementação

A Figura 11 apresenta a interface gráfica do protótipo, onde pode-se observar os seguintes componentes:

- a) botão para selecionar e carregar um arquivo texto, sendo que, caso o arquivo texto selecionado tenha sido gerado através do protótipo e tenha sido traduzido, a tradução para LIBRAS também será carregada;
- b) botão para salvar o texto de entrada (campo E) em um arquivo texto, no formato txt, sendo que, caso tenha sido anteriormente traduzido, um arquivo auxiliar com a tradução correspondente para LIBRAS também será salvo;
- c) botão para traduzir o texto de entrada (campo E) para LIBRAS;
- d) botão para limpar o texto de entrada (campo E) e a tradução correspondente (campo F);
- e) campo para o texto de entrada que será traduzido;
- f) campo para a transcrição para LIBRAS do texto de entrada, sendo que durante o processo de tradução, o texto é dividido em sentenças e a primeira sentença é mostrada nesse campo;
- g) campo para contabilizar o total de sentenças contidas no texto de entrada, após realizado o processo de tradução;
- h) campo para mostrar em que sentença está a reprodução dos vídeos;
- i) campo para, caso o usuário queira, adicionar um período de pausa (mensurado em

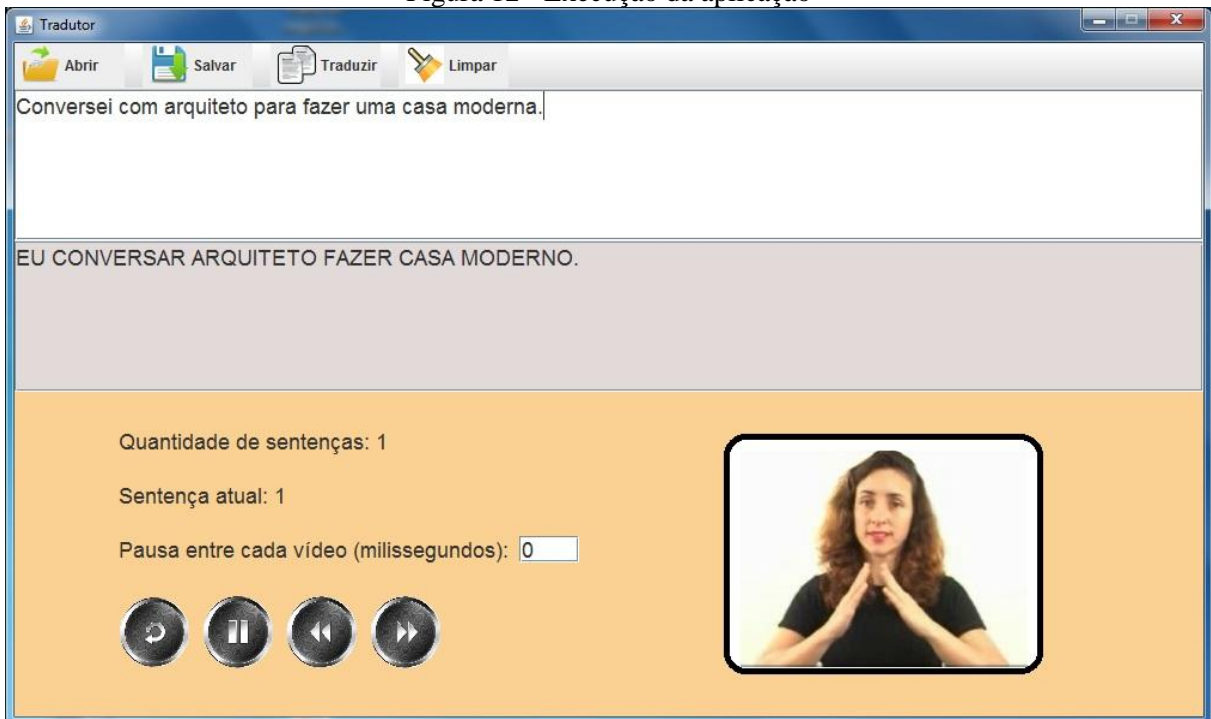
- milissegundos) entre a exibição de cada sinal;
- j) botão para iniciar a reprodução dos sinais de LIBRAS. Quando estiver reproduzindo, se pressionado novamente, reinicia a exibição da sentença. No momento que acabar a exibição de uma sentença, caso tenha outra, a exibição dela é iniciada automaticamente. Se não existir outra sentença, a exibição é encerrada;
 - k) botão para, caso pressionado durante a reprodução dos vídeos, pausar a exibição. Se pressionado novamente, continua a exibição a partir do sinal que parou;
 - l) botão para voltar para uma sentença anterior, caso exista. É possível executar essa ação tanto quando os sinais estiverem sendo exibidos, como quando a reprodução estiver parada;
 - m) botão para avançar uma sentença, caso exista. É possível executar essa ação tanto quando os sinais estiverem sendo exibidos, como quando a reprodução estiver parada;
 - n) campo para exibir os vídeos dos sinais de LIBRAS.

Figura 11 - Interface gráfica da aplicação



De forma resumida, o processo principal que o usuário efetuará é o seguinte: digitar ou carregar um texto de entrada, pressionar o botão para traduzir, aguardar a tradução ser concluída e pressionar o botão para reproduzir os vídeos com os sinais de LIBRAS. A Figura 12 mostra a tradução da sentença “Conversei com arquiteto para fazer uma casa moderna.” para LIBRAS.

Figura 12 - Execução da aplicação



3.4 RESULTADOS E DISCUSSÕES

Com o intuito de validar o que estava sendo desenvolvido, bem como para tirar algumas dúvidas em relação a gramática de LIBRAS, o protótipo foi apresentado para a professora de LIBRAS da FURB, Fabiana Schmitt Corrêa. Foram realizados dois encontros, nos dias 23 de abril de 2015 e 20 de maio de 2015.

No primeiro encontro, foi apresentada a ideia do trabalho e sanadas algumas dúvidas. No segundo encontro, o protótipo foi apresentado e avaliado pela professora Fabiana, que teve impressões positivas da ferramenta, porém, levantou algumas ressalvas, sobretudo relacionadas à limitação dos sinais de saída, incluindo:

- a) em LIBRAS, como abordado na seção 2.1.3 da fundamentação teórica, cada tipo de frase possui uma expressão facial e corporal. O Dicionário da Língua Brasileira de Sinais traz os sinais na forma afirmativa. Assim, para frases interrogativas, convencionou-se adicionar o sinal perguntar antes da frase, porém, quando uma frase for do tipo exclamativa, o protótipo não fará distinção;
- b) muitos sinais não estão presentes no Dicionário da Língua Brasileira de Sinais, sendo extremamente necessário trabalhar com sinônimos, pois em textos longos, uma grande quantidade de palavras são soletradas.

Além disso, observa-se que as seguintes regras do sistema de transcrição não estão em total acordo com o utilizado pela FENEIS:

- a) duas ou mais palavras da Língua Portuguesa representadas por um único sinal são escritas separadas por hífen, como em SABER-NÃO, CORTAR-CABELO. No protótipo não estão representados com hífen em todos os casos, pois o Dicionário da Língua Brasileira de Sinais nem sempre traz esses sinais com hífen;
- b) palavras da Língua Portuguesa representadas por mais de um sinal têm as palavras que representam cada sinal separadas pelo símbolo ^, como, por exemplo, a palavra zebra, é representada como CAVALO^LISTRA. No entanto, no Dicionário da Língua Brasileira de Sinais, a palavra zebra está colocada como um único vídeo, que mostra os dois sinais em sequência. Logo, no protótipo aparecerá apenas a palavra ZEBRA. Assim sendo, casos em que a palavra está no Dicionário da Língua Brasileira de Sinais, é necessário adicionar uma regra no dicionário de exceções;
- c) não há distinção de gênero e número em LIBRAS, ou seja, na representação das palavras que contêm gênero e número, a letra que distingue o gênero é substituída pelo símbolo @, como EL@ para representar ele, ela, eles e elas. No protótipo, palavras que entram nessa regra são sempre representadas no masculino e no singular. Para as palavras ele, ela, eles e elas aparecerá o símbolo ELE.

Por fim, nota-se que por motivo de simplificação do desenvolvimento, para palavras que possuem mais de um sinal, o tradutor sempre mostrará o primeiro sinal, independente do contexto da sentença. O Quadro 14 mostra exemplos de frases com a palavra andar, que dependendo o contexto, ocorrem problemas na exibição dos sinais.

Quadro 14 – Exemplo de frases em português com a palavra andar

Frase 1	O elevador parou no 2º andar.
Frase 2	Ele andava rápido.
Frase 3	O lobo anda camuflado na floresta.

Na primeira frase do Quadro 14, a palavra andar é classificada como substantivo, referente ao piso de edifícios, enquanto nas demais frases é classificada como verbo, significando o ato de caminhar. No entanto, em todos os casos, o sinal exibido pelo protótipo será o do substantivo, tornando-se errado na demais situações. Ainda levando-se em conta o caso específico dessa palavra, observa-se que quando classificada como verbo, possui sinais diferentes para quando a ação é relativo à pessoa, animal ou objeto. Ou seja, para cada frase do Quadro 14, existe um sinal em LIBRAS diferente para andar.

Em relação aos trabalhos correlatos, o Quadro 15 mostra comparações de características com o protótipo desenvolvido.

Quadro 15 - Comparativo entre trabalhos correlatos e o protótipo desenvolvido

características	protótipo desenvolvido	ProDeaf	Poli-Libras	Librol
plataforma	<i>desktop</i>	móvel e <i>web</i>	<i>web</i>	<i>desktop</i>
entrada por texto	sim	sim	sim	sim
reconhecimento de voz	não	sim	não	não
sinais com vídeos	sim	não	não	não
sinais com animação	não	sim	sim	não
exibição da transcrição do texto em português para LIBRAS	sim	não	não	sim

4 CONCLUSÕES

Visto a quantidade de pessoas que utilizam LIBRAS, a dificuldade muitas vezes encontrada por pessoas que não dominam essa língua conseguir se comunicar com seus usuários e mesmo a falta de ferramentas que possibilitem uma maior interação em seu aprendizado, decidiu-se elaborar uma ferramenta de simples utilização que auxilie na compreensão de LIBRAS.

Durante a construção do protótipo, encontrou-se algumas limitações das tecnologias adotadas: (i) o Dicionário da Língua Brasileira de Sinais, ainda que tenha uma grande quantidade de sinais e possua a maioria dos que são recorrentes em LIBRAS, não contempla todos os sinais difundidos, sendo que todos se encontram na forma afirmativa; (ii) a biblioteca Cogroo, utilizada para a realização da etiquetagem de palavras, possui algumas falhas. Para amenizar esses problemas, criou-se o dicionário de exceções, que consegue tratar parte desses casos, sendo necessário criar as regras que devem ser levadas em conta na análise. Também não são tratados casos de palavras com mais de um sinal, pois exige um tratamento que não foi possível implementar durante o desenvolvimento do trabalho.

Deve-se ressaltar que, mesmo com algumas falhas, a biblioteca Cogroo mostrou-se a melhor alternativa disponível para realizar a etiquetagem de palavras em Java. O mesmo é válido para o Dicionário da Língua Brasileira de Sinais, que diante do proposto, é a tecnologia mais adequada. Observa-se que, futuramente, caso queira-se adicionar novos sinais ou mesmo trocar toda a saída do protótipo, não haverá necessidade de modificar a implementação.

Diante dos testes realizados, comprovou-se que o protótipo atingiu os objetivos propostos, com algumas limitações, como abordado anteriormente, sendo o caso de palavras com mais de um sinal a maior falha de tradução. Tendo esse cenário em vista, nesse estágio de desenvolvimento, o protótipo mostra-se mais adequado para ser usado com a supervisão de um especialista em LIBRAS, podendo servir como complemento para o ensino e compreensão da língua, pois tal especialista poderá determinar quais sentenças estão traduzidas corretas e apontar quais eventualmente possuem alguma falha.

4.1 EXTENSÕES

Como sugestão de melhorias para o protótipo propõe-se:

- a) refinar o processo de tradução, criando um analisador de contexto do texto de entrada, para nos casos de palavras com mais de um sinal, conseguir distinguir qual é o sinal correto em cada situação;

- b) elaborar novas formas de entrada de texto, utilizando reconhecimento de voz ou lendo outros formatos de arquivo de texto;
- c) desenvolver uma nova interface de saída, utilizando computação gráfica, possibilitando que um especialista crie sinais, para que a ferramenta não fique limitada aos vídeos do Dicionário da Língua Brasileira de Sinais;
- d) adaptar a ferramenta para dispositivos móveis, aumentando os potenciais usuários finais.

SILVA, Bento C. D. et al. **Introdução ao processamento das línguas naturais e algumas aplicações**. São Carlos, 2007. Disponível em:
<<http://www.icmc.usp.br/~tasparado/NILCTR0710-DiasDaSilvaEtAl.pdf>>. Acesso em: 25 mai. 2015.