

CENTRO FEDERAL DE EDUCAÇÃO TECNOLÓGICA DE MINAS GERAIS –
CAMPUS TIMÓTEO

Engenharia de Computação

Jéssica Oliveira Neves

MÁQUINA DE TRADUÇÃO AUTOMÁTICA PORTUGUÊS - LIBRAS

Timóteo

2014

Jéssica Oliveira Neves

MÁQUINA DE TRADUÇÃO AUTOMÁTICA PORTUGUÊS - LIBRAS

Monografia apresentada ao Curso de Engenharia de Computação do Centro Federal de Educação Tecnológica de Minas Gerais, como requisito parcial para obtenção do título de Engenheira de Computação.

Orientador: Aléssio Miranda Júnior

Timóteo

2014

Jéssica Oliveira Neves

MÁQUINA DE TRADUÇÃO AUTOMÁTICA PORTUGUÊS - LIBRAS

Monografia apresentada ao Curso de Engenharia de Computação do Centro Federal de Educação Tecnológica de Minas Gerais, como requisito parcial para obtenção do título de Engenheira de Computação.

Aléssio Miranda Júnior (Orientador) - CEFET/MG

Deisymar Botega Tavares - CEFET/MG

Convidado 1 - Instituição

Timóteo, 13 de março de 2014.

Dedico esta monografia a minha família pela fé e confiança demonstrada.

Aos meus amigos, em especial ao Moisés, pelo apoio incondicional.

Aos professores pelo simples fato de estarem dispostos a ensinar.

Ao orientador pela paciência demonstrada no decorrer do trabalho.

Enfim a todos que de alguma forma tornaram este caminho mais fácil de ser percorrido.

AGRADECIMENTOS

Agradeço a Deus pela oportunidade de estar realizando este trabalho.

A minha família, pelo incentivo e colaboração, principalmente nos momentos de dificuldade.

Ao meu orientador por estar sempre disposto a ajudar.

Agradeço aos meus colegas, em especial ao Moisés, pelas palavras amigas nas horas difíceis, pelo auxílio nos trabalhos e dificuldades e principalmente por estarem comigo nesta caminhada tornando-a mais fácil e agradável.

Que os vossos esforços desafiem as impossibilidades, lembrai-vos de que as grandes coisas do homem foram conquistadas do que parecia impossível. (CHAPLIN)

RESUMO

De acordo com a LEI Nº 10.436 (2002), deve ser garantido formas institucionalizadas de apoiar o uso e difusão da Língua Brasileira de Sinais (LIBRAS) como meio de comunicação objetiva e de utilização corrente das comunidades surdas do Brasil. A LIBRAS é o meio no qual o surdo entra em contato com o ambiente ao seu redor, porém ela não substitui a modalidade escrita da língua portuguesa. Cerca de 60% da população surda brasileira conhece a LIBRAS, mas mesmo assim, a comunicação com os ouvintes é difícil. Isso se deve ao fato de que, a maioria da população ouvinte, desconhece a língua de sinais o que dificulta o processo de integração. Tendo em vista essa dificuldade do estabelecimento da comunicação, percebe-se a necessidade da criação de mecanismos para auxiliar esse processo. A Tradução de Línguas de forma Automática seria uma possível solução para esse problema. Esse trabalho, portanto, tem como objetivo principal a aplicação do Apertium, uma Máquina de Tradução Automática baseada em regras sintáticas superficiais, que utiliza uma base de conhecimento própria, com uma estrutura aberta e flexível no padrão XML, como ferramenta de tradução de línguas, utilizando como caso especial o par de tradução Português - LIBRAS.

Palavras-chave: Tradução automática. Máquinas de tradução automática. LIBRAS. Linguística computacional. Regras de tradução. Apertium.

ABSTRACT

According to Law No. 10,436 (2002), must be guaranteed institutionalized forms of support the use and dissemination of the Brazilian Sign Language (LIBRAS) as a communication objective and current use of deaf communities in Brazil. The LIBRAS is the way in which the deaf comes into contact with the environment around them, but it does not replace the written form of the Portuguese language. About 60% of the deaf Brazilian population knows LIBRAS, but even so, the communication with the audience is difficult. This is due to the fact that most of the population listener don't know the sign language which makes the integration process to be hard. Given the difficulty of establishing communication, one realizes the need for mechanisms to assist this process. The translation of languages so Auto would be a possible solution to this problem. This work, therefore, has as main objective the implementation of Apertium, a Machine Translation Automatic rule-based surface syntactic, that uses a knowledge base of their own, with an open and flexible XML standard, as a tool for language translation using as a special case the translation pair Portuguese - LIBRAS.

Keywords: Machine translation. Automatic translation machines. LIBRAS. Computational linguistics. Translation rules. Apertium.

LISTA DE FIGURAS

Figura 1 - Ideia Inicial do Projeto Falibras.	19
Figura 2 - Sinal que representa a palavra Computador.	20
Figura 3 - Exemplo de Pronomes Pessoais	23
Figura 4 - Exemplo Advérbios de Tempo	25
Figura 5 - Verbo IR.	26
Figura 6 - A tradução por máquina no panorama atual.	29
Figura 7 - Arquitetura simplificada do Apertium.	30
Figura 8 - Processo de Tradução da Frase.	45
Figura 9 - Tradução no Apertium-viewer.	49

LISTA DE QUADROS

Quadro 1 - Exemplos de Símbolos utilizados no Apertium	37
Quadro 2 - Arquivos utilizados na tradução Português-LIBRAS	38

LISTA DE CÓDIGOS

Código 1 - Definição base de dicionários	40
Código 2 - Alfabeto do dicionário monoligüe de LIBRAS	40
Código 3 - Exemplo de classificação dos símbolos para a LIBRAS.....	41
Código 4 - Paradigma para as palavras EL@ e EL@s	42
Código 5 - Exemplo <section> dicionário LIBRAS.	42
Código 6 - Exemplo do mapeamento de palavras Português - LIBRAS	43
Código 7 - Estrutura básica do arquivo de Regras de Transferência	44
Código 8 - Exemplo definição de Categoria e Atributo.....	46
Código 9 - Substituição de Símbolos.	47
Código 10 - Inserção do Advérbio de Tempo.....	47
Código 11 - Inclusão do Pronome Pessoal	48

LISTA DE ABREVIATURAS

Apertium - Opentrad Apertium Machine Translation System

HTML - Hypertext Markup Language

INES - Instituto Nacional da Educação de Surdos

LIBRAS - Língua Brasileira de Sinais

MTA - Máquinas de Tradução Automática

RTF - Rich Text Format

TA - Tradução Automática

XML - Extensible Markup Language

SUMÁRIO

1. INTRODUÇÃO	14
1.1. Definição do Problema	14
1.2. Motivação.....	15
1.3. Objetivo	16
2. ESTADO DA ARTE	17
2.1. Desafios da tradução Português - LIBRAS.....	18
2.2. Falibras	19
2.3. InfoLIBRAS	20
3. FUNDAMENTOS TEÓRICOS.....	21
3.1. Linguagem Natural.....	21
3.2. Linguagem Brasileira de Sinais	22
3.2.1. <i>Pronomes Pessoais</i>	22
3.2.2. <i>Substantivos</i>	23
3.2.3. <i>Advérbios de Tempo</i>	23
3.2.4. <i>Verbos</i>	25
3.3. Linguística Computacional	26
3.4. Tradução Automática.....	27
3.4.1. <i>Máquina de Tradução Automática</i>	28
3.5. Apertium	29
4. METODOLOGIA	33
5. RESULTADOS	35
5.1. Conceitos básicos relacionados ao Apertium.....	36
5.2. Desenvolvendo um novo par de línguas no Apertium.....	37
5.3. Dicionários Monolíngues.....	39
5.4. Dicionário Bilíngue.....	43
5.5. Regras de Tradução.....	43
5.5.1. <i>Criando Regras de Tradução no Apertium</i>	46
5.6. Considerações Finais	48

6. CONCLUSÕES E PROPOSTAS DE CONTINUIDADE.....	50
REFERÊNCIAS.....	52
APÊNDICE A - DICIONÁRIO MONOLÍNGUE LIBRAS.....	54
APÊNDICE B - DICIONÁRIO BILÍNGUE PORTUGUÊS - LIBRAS	56
APÊNDICE C - REGRAS DE TRANSFERÊNCIA PORTUGUÊS - LIBRAS	58

1. INTRODUÇÃO

Este trabalho pressupõe que o aprendizado da Língua Brasileira de Sinais (LIBRAS) é de fundamental importância para surdos e ouvintes uma vez que interfere diretamente na obtenção de conhecimentos por parte daqueles que possuem algum tipo de deficiência auditiva.

1.1. Definição do Problema

De acordo com os dados do Censo Demográfico (IBGE, 2010), aproximadamente 7.574.145 brasileiros são portadores de algum tipo de deficiência auditiva e cerca de 344.206 não conseguem ouvir de modo algum, ou seja são surdos. Desses, cerca de 60% conhecem a LIBRAS, mas mesmo assim, a comunicação com os ouvintes é difícil. Isso se deve ao fato de que, a maioria da população ouvinte, desconhece a língua de sinais o que dificulta o processo de integração, e principalmente, o processo de educação dos surdos.

Tendo em vista essa dificuldade do estabelecimento da comunicação, percebe-se a necessidade da criação de mecanismos para auxiliar esse processo. A Tradução de Línguas de forma Automática é uma possível solução para esse problema. Consiste em traduzir automaticamente um texto escrito em uma língua natural (chamada de língua-origem) para outra língua natural (língua-destino), traduzindo o conteúdo de maneira a produzir um resultado inteligível e que preserve suas características como estilo, coesão e significado.

Sabe-se que as mais bem sucedidas Máquinas de Tradução Automática (MTA) são baseadas em softwares proprietários, fechados e com propósito comercial o que dificulta o desenvolvimento de novos pares de tradução por não ser possível o estudo de forma sistemática da ferramenta. É definido como "par de tradução" o conhecimento necessário para que um tradutor transforme um texto em uma determinada língua-origem para uma outra língua-destino (MIRANDA JR., 2009).

Acredita-se que transformar essa tecnologia em algo aberto e democrático traz resultados positivos uma vez que possibilita o desenvolvimento da comunidade.

Dentre várias MTAs que surgiram seguindo esse modelo de software livre, destaca-se o Opentrad Apertium (Apertium).

O Apertium é uma MTA baseada em regras sintáticas superficiais, que utiliza uma base de conhecimento própria, com uma estrutura aberta e flexível no padrão XML (MIRANDA JR., 2009).

A ausência de uma interface específica para a manutenção da base de conhecimentos linguísticos é um obstáculo para a popularização do Apertium e reduz o número de colaboradores potenciais.

1.2. Motivação

De acordo com a LEI Nº 10.436 (2002), deve ser garantido formas institucionalizadas de apoiar o uso e difusão da Língua Brasileira de Sinais - LIBRAS como meio de comunicação objetiva e de utilização corrente das comunidades surdas do Brasil.

A LIBRAS é o meio no qual o surdo entra em contato com o ambiente ao seu redor, porém ela não substitui a modalidade escrita da língua portuguesa. Uma vez confirmada a importância do conhecimento da LIBRAS por parte de surdos e ouvintes vê-se a necessidade da criação de ferramentas que facilitem esse aprendizado.

Um dos benefícios de um tradutor geral Português - LIBRAS é a hipótese de que a visualização da mesma mensagem na língua portuguesa e na LIBRAS facilitará o aprendizado tanto do Português, por parte dos usuários surdos, como da LIBRAS, por parte dos usuários ouvintes (FRANCO, SILVA, & BRITO, [201-?]).

Esse trabalho contribuirá para a inclusão do surdo na sociedade e nos meios digitais. Tendo em vista que a inclusão digital além de facilitar o processo de aprendizagem do surdo, também ajuda na construção do processo de conceituação e no desenvolvimento de habilidades importantes para que ele participe da sociedade do conhecimento.

1.3. Objetivo

Uma vez que os deficientes auditivos representam uma boa parcela da população brasileira atual, existe a necessidade de criação de mecanismos que facilitem o estabelecimento da comunicação entre surdos e ouvintes. Principalmente, mecanismos que possam ser acessíveis à todo tipo de usuário, não só especialistas em Computação e Linguística, como também à possíveis usuários leigos.

A língua de sinais é importante para a educação do surdo, pois é a sua principal forma de comunicação. Através dos sinais o surdo pode compreender com mais facilidade o mundo em que vive.

Entende-se por LIBRAS a forma de comunicação e expressão, em que o sistema linguístico é de forma visual-motora. Possui uma estrutura gramatical própria, que constitui um sistema linguístico de transmissão de ideias e fatos, oriundos de comunidades de pessoas surdas no Brasil (FAQUETI C. , GRANDI, FANTINI, & LORENZETTI, 2005).

Tendo em vista esse cenário, esse trabalho tem como objetivo principal a aplicação do Apertium como ferramenta de tradução de línguas, utilizando como caso especial o par de tradução Português - LIBRAS.

Prevê-se também a criação de uma pequena especificação Português - LIBRAS para facilitar o desenvolvimento da tradução de ambas as línguas.

Para esse objetivo ser alcançado é necessário além de tudo, o estudo sistemático do Apertium, bem como todo fundamento da Linguística Computacional e das Máquinas de Tradução.

2. ESTADO DA ARTE

O Instituto dos Surdos-Mudos, hoje, Instituto Nacional da Educação de Surdos (INES) foi a primeira escola para surdos no Brasil, fundada em 1857. Foi a partir deste, com a miscigenação da antiga língua de sinais brasileira com a Língua de Sinais Francesa, que nasceu, definitivamente, a Língua Brasileira de Sinais (LÍNGUA BRASILEIRA DE SINAIS, 2013).

Através de diversos movimentos e muita pesquisa na área, foi legitimada como Língua a comunicação gestual entre os surdos. E, no fim do século XX, os movimentos se intensificaram para a oficialização da LIBRAS no país, criando-se em 1993 o primeiro projeto de Lei.

Porém, somente em 24 de abril de 2002, através da Lei 10.436, a LIBRAS foi oficialmente aceita e reconhecida como a segunda língua oficial brasileira. Em 2006, a LIBRAS foi regulamentada como disciplina curricular através do Decreto 5.626. E, em 2007, ela foi estruturada Língua por possuir complexidade própria e comunicação eficaz. Finalmente em 2010, foi regulamentada a profissão de Tradutor/Interprete de LIBRAS.

Todo indivíduo é capaz de desenvolver o aprendizado, então, quando é exposto a uma língua, começa o processo de aprendizado com uma cadeia de significados que é a base para o desenvolvimento da escrita (PEREIRA, [20--]). Com os surdos não é diferente. Apesar de a surdez dificultar a comunicação, muitos métodos e tentativas educacionais estão sendo estudadas e discutidas para a educação, principalmente das crianças, surdas.

Com a LIBRAS, o processo de comunicação entre surdos e ouvintes é facilitado. Por exemplo, para um educador que possui um aluno surdo em sala de aula, a língua de sinais é de fundamental importância para a transmissão do conteúdo e, conseqüentemente, do conhecimento.

A realização da tradução automática entre o Português e a LIBRAS facilitaria todo esse processo de integração, sabe-se, porém, que esta não é uma tarefa fácil visto que existem diversos desafios da tradução do Português para a LIBRAS.

2.1. Desafios da tradução Português - LIBRAS

No início do processo de descrição gramatical de textos em LIBRAS para o processamento de linguagem natural deparou-se com expressões linguísticas que não poderiam ser classificadas em termos de classes gramaticais assim, tentou-se identificar essas sequencias de expressões que se referiam a uma mesma entidade.

Questões básicas começaram a ser discutidas como por exemplo, quando e como as relações são estabelecidas entre os componentes de uma frase. Deparou-se então, com a impossibilidade de traduzir uma língua, que se apresenta tridimensionalmente, a partir de unidades lineares.

Com o passar dos anos, a abordagem do estudo das línguas foi sendo modificada passando de apenas uma manifestação biológica da natureza humana para uma rede constituída de elementos coesivos que estabelecem coerência.

O advento da Linguística Computacional, que utiliza procedimentos automáticos para análise de dados, tem expandido o conceito de classe de palavras para o de categorias morfossintáticas, a fim de melhor aproveitar as informações disponíveis nas formas literais. Para isso já existem programas capazes de atribuir, automaticamente, classes gramaticais às palavras de um texto. Estes programas são geralmente chamados de etiquetadores de estruturas/categorias morfossintáticas, cuja complexidade do esquema de rótulos tem variado, mas sempre têm estado presentes certas classificações básicas como substantivo, verbo, adjetivo e pronome. Mas embora esses etiquetadores sejam capazes de atribuir uma classificação a todas as palavras de um texto, há necessidade de se repensar essas categorias para contemplar especificidades das línguas gesto-visuais (FELIPE, 2007a).

Os primeiros estudos, as línguas de sinais, mostravam que as línguas de sinais possuem os mesmos universos linguísticos com relação às línguas orais auditivas, variando apenas a modalidade. Mas, como essas línguas possuem uma tridimensionalidade e outras características que as línguas oral-auditivas não possuem, outros paradigmas precisaram ser criados.

Assim, começaram as análises dos elementos não manuais, ou seja, os estudos das expressões faciais e corporais.

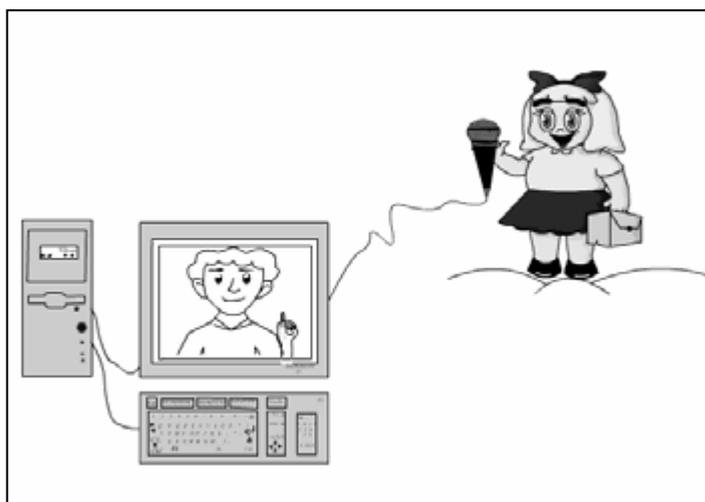
Vários projetos visando a inclusão social e digital dos surdos e o processo de tradução automática entre as duas línguas, Português e LIBRAS, estão sendo criados e discutidos no Brasil. Dentre esses projetos, pode-se citar o Falibras e o InfoLIBRAS que estão detalhados nas seções a seguir.

2.2. Falibras

Com o Falibras surgiu com o objetivo de facilitar a transmissão de conhecimentos, entre educadores e alunos com deficiências auditivas ou surdos. Desde 2001 o professor do Instituto de Computação da Universidade Federal de Alagoas, Patrick Henrique Brito, vem trabalhando com o objetivo de promover a inclusão social de deficientes auditivos, que têm dificuldades de aprender a língua portuguesa e que se comunicam apenas pela linguagem de sinais.

De uma maneira geral, o sistema Falibras capta a voz e a transforma em texto usando recursos do IBM Viavoice. Em seguida, esse texto é analisado por um interpretador que, além de corrigir a ortografia, definirá o contexto em que as palavras estão inseridas na frase, a fim de mostrar a tradução adequada em LIBRAS (CORADINE, ALBURQUEQUE, SILVA, MADEIRO, & PEREIRA, 2004). A Figura 1 ilustra a ideia principal do sistema Falibras.

Figura 1 - Ideia Inicial do Projeto Falibras.



Fonte: (CORADINE, ALBURQUEQUE, SILVA, MADEIRO, & PEREIRA, 2004)

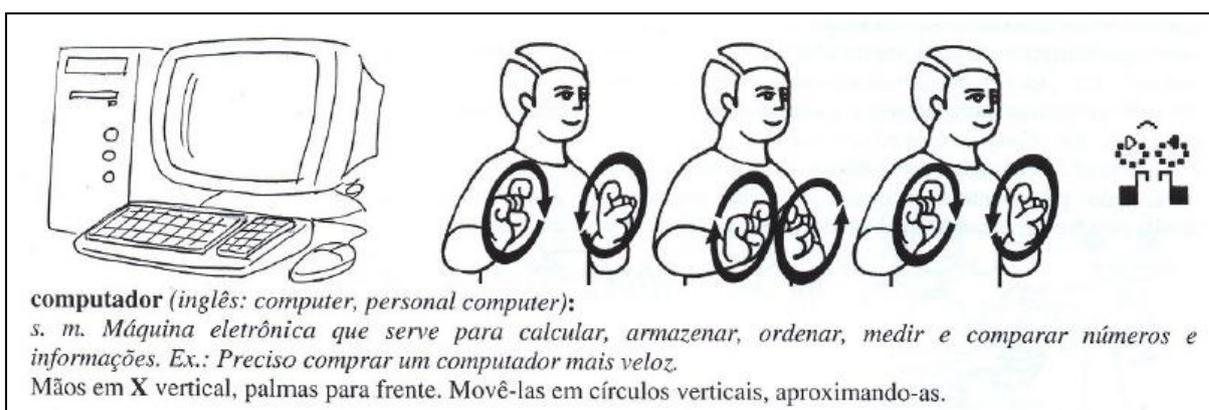
Atualmente, o Falibras juntamente com uma empresa alagoana que investiu no projeto desenvolveram um aplicativo, Hand Talk (da tradução em português “Mãos que Falam”), para celular que venceu o prêmio de melhor projeto na categoria de Inclusão Social no concurso WSA-Mobile promovido pela ONU. A expectativa é que até 2014 seja lançada a primeira parte do projeto, voltada para dispositivos móveis, sendo disponibilizada na Apple Store e na Google Play.

2.3. InfoLIBRAS

Outra ferramenta com o objetivo de facilitar a comunicação entre surdos e ouvintes foi proposta em 2005, no XXV Congresso da Sociedade Brasileira de Computação, e ficou conhecida como InfoLIBRAS (Informática em Língua Brasileira de Sinais).

O InfoLIBRAS é um ambiente para auxiliar o ensino da LIBRAS através da Internet. O sistema foi criado com o intuito de complementar o Dicionário da Língua Portuguesa de Sinais (LIBRAS) pois este não contempla totalmente as palavras relacionadas com a informática. De acordo com os autores, (FAQUETI C. , GRANDI, FANTINI, & LORENZETTI, 2005), o sistema será útil para novos alunos surdos de cursos de informática. Para tanto, os sinais que designam as palavras precisam ser criados através de movimentos corporais. Como no exemplo mostrado pela Figura 2.

Figura 2 - Sinal que representa a palavra Computador.



Fonte: (CAPOVILLA & RAPHAEL, 2001)

3. FUNDAMENTOS TEÓRICOS

O processo de comunicação e de troca de informação para o ser humano é algo natural, seja por meio da fala e da escrita ou através de gestos e sinais.

3.1. Linguagem Natural

Entende-se por linguagem natural ou linguagem o meio utilizado para a comunicação entre os seres humanos que é continuamente estabelecido e aperfeiçoado ao longo do processo construtivo e histórico da sociedade. As linguagens naturais estão em constante mutação, adaptação, criação e extinção. A fala e a língua de sinais são exemplos de linguagens naturais (FERNANDES, 2002).

Sabe-se que a fala é constituída por repetição de palavras, cujo sentido se repete, e estas por sua vez são compostas por vogais e consoantes que também se repetem. As línguas faladas variam no tempo e no espaço, ou seja, variam de acordo com a cultura do local onde é falada. O Português do Brasil, por exemplo, é diferente do Português de Portugal.

As línguas de sinais por sua vez, são altamente expressivas utilizando expressões faciais e corporais, em vez de sons na comunicação. As línguas de sinais são as línguas naturais das comunidades de surdos utilizadas para a troca de informações entre ouvintes, surdos e deficientes auditivos.

Atualmente, com a globalização, é cada vez mais necessário a troca de informações e conhecimentos entre os povos porém esse processo é dificultado pela existência das inúmeras línguas, cada qual com suas particularidades. Vê-se então a necessidade da criação de intérpretores e/ou tradutores para o estabelecimento da comunicação.

O grande desafio dos estudiosos da área é sistematizar e estabelecer precisamente as regras de construção de uma linguagem natural. Este processo é um desafio para a área da Ciência da Computação e ainda não foi automatizado de maneira satisfatória e sua execução se enquadra em uma categoria de problemas considerados computacionalmente muito complexos (HUTCHINS & SOMERS, 1992).

A criação de regras para a tradução automática é um meio de estabelecer o relacionamento entre línguas.

3.2. Linguagem Brasileira de Sinais

A Língua Brasileira de Sinais (LIBRAS) é uma das línguas faladas no Brasil. Se trata de uma língua gestual destinada, principalmente, às comunidades surdas.

O parágrafo 2º do artigo 12 da Resolução do CNE/CEB nº 2/2001, que instituiu as Diretrizes Nacionais para a Educação Especial na Educação Básica, definiu que ao aluno surdo deve ser assegurado o acesso aos conteúdos curriculares, mediante a utilização de língua de sinais, sem prejuízo do aprendizado da língua portuguesa (FELIPE, 2007b).

A LIBRAS é, sobretudo, uma linguagem natural que vem sendo continuamente estabelecida e aperfeiçoada ao longo do processo construtivo e histórico. Ela possui regras gramaticais próprias que podem divergir, e muito, da língua portuguesa.

Esta seção tem como objetivo apresentar algumas classes gramaticais da LIBRAS, e suas particularidades, que são utilizadas ao longo deste trabalho.

3.2.1. Pronomes Pessoais

A LIBRAS possui um sistema pronominal para representar as pessoas do discurso. Que podem ser divididos em:

- Primeira pessoa (singular, dual, trial, quatrial e plural), sendo esses pronomes, respectivamente, EU, NÓS-2, NÓS-3, NÓS-4, NÓS-GRUPO, NÓS/NÓS TOD@S.
- Segunda pessoa (singular, dual, trial, quatrial e plural), sendo esses pronomes, respectivamente, VOCÊ, VOCÊ-2, VOCÊ-3, VOCÊ-4, VOCÊ-GRUPO, VOCÊS/VOCÊS TOD@S.
- Terceira pessoa (singular, dual, trial, quatrial e plural), sendo esses pronomes, respectivamente, EL@, EL@-2, EL@-3, EL@-4, EL@S-GRUPO, EL@S/EL@S TOD@S. Diferentemente do Português, os pronomes pessoais

na terceira pessoa não possuem marca para gênero (masculino e feminino) e o símbolo que é utilizado para marcar essa ausência de gênero é o "@".

Na Figura 3 têm-se a representação de alguns desses pronomes pessoais.

Figura 3 - Exemplo de Pronomes Pessoais



Fonte: (FELIPE, 2007b)

3.2.2. Substantivos

É interessante observar que não há flexão de gênero em LIBRAS, os substantivos e adjetivos são, em geral, não marcados. Entretanto, quando se quer explicitar substantivos dentro de determinados contextos, a indicação de sexo é feita adicionando-se o sinal "HOMEM / MULHER", para pessoas e animais, ou a indicação é obtida através de sinais diferentes para um e para outro sexo.

Exemplos:

HOMEM "homem"

HOMEM^VELH@ "vovô"

MULHER "mulher"

MULHER^VELH@ "vovó"

3.2.3. Advérbios de Tempo

O tempo, na LIBRAS, é marcado sintaticamente através de advérbios de tempo que indicam se a ação está:

- No presente: HOJE, AGORA;
- Ocorreu no passado: ONTEM, ANTEONTEM;
- Irá ocorrer no futuro: AMANHÃ.

Os advérbios de tempo geralmente vêm especificados no começo da frase, mas podem ser usados também no final. Quando não há, na frase, um advérbio de tempo específico, geralmente a frase está no presente, e não há nenhuma especificação temporal.

A Figura 4 exemplifica alguns dos advérbios de tempo.

Figura 4 - Exemplo Advérbios de Tempo



Fonte: (FELIPE, 2007b)

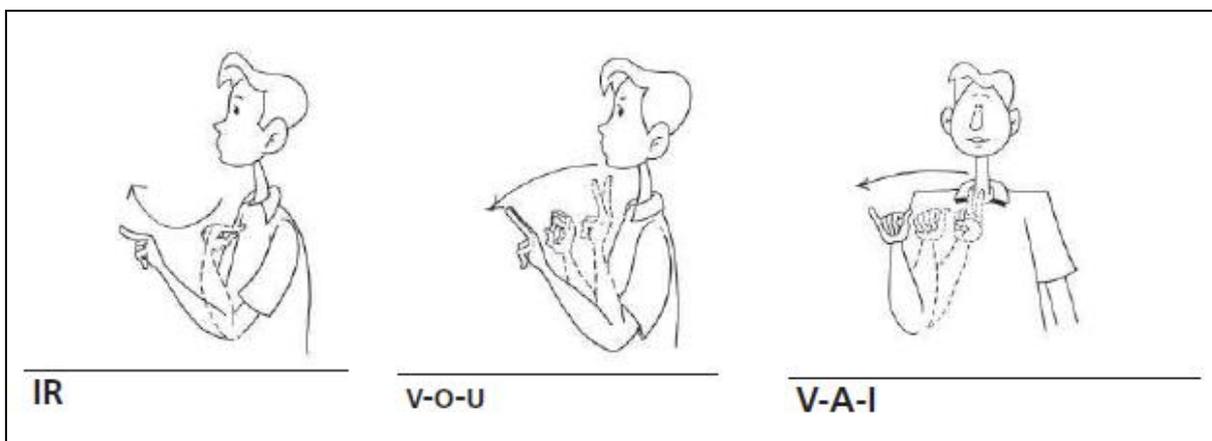
3.2.4. Verbos

Na LIBRAS não há marca de tempo nas formas verbais, nas frases, muitos verbos ficam no infinitivo e são acompanhados de algum advérbio de tempo. Pode-se dizer que existem três tipos de verbos:

- Verbos simples, ou aqueles que não flexionam.
- Verbos com concordância, são os verbos que flexionam.
- Verbos especiais, são os verbos que têm afixos locativos, como por exemplo os verbos VIR e IR. Na Libras, o verbo "IR" possui uma forma neutra, como a maioria dos verbos, mas possui também formas que marcam flexões

peçoais que podem ser empréstimos da forma verbal em português. Conforme a Figura 5.

Figura 5 - Verbo IR



Fonte: (FELIPE, 2007b)

3.3. Linguística Computacional

A área responsável pela investigação do tratamento computacional da linguagem e das línguas naturais é conhecida como Linguística Computacional.

A Linguística Computacional é a área de conhecimento que explora as relações entre linguística e informática, tornando possível a construção de sistemas com capacidade de reconhecer e produzir informação apresentada em linguagem natural (VIEIRA & LIMA, 2001).

Todos os estudos relacionados a Linguística Computacional é utilizado para tentar processar as línguas naturais e dominar o conhecimento linguístico envolvido no domínio de uma linguagem natural. Para tanto, a Linguística Computacional se divide em duas subáreas: a Linguística de Corpus e o Processamento de Linguagem Natural (PLN).

A Linguística de Corpus se ocupa da coleta e análise de corpus, que é um conjunto de dados linguísticos coletados criteriosamente para serem objeto de pesquisa linguística. Essas amostras podem ser de diferentes fontes. Por isso, pode-se encontrar os mais variados bancos de corpus eletrônicos: há corpus de

linguagem falada, corpus de linguagem escrita literária, corpus com textos de jornal, corpus compostos exclusivamente por falas de crianças em estágio de desenvolvimento linguístico, etc (OTHERO, 2006).

Já a área de Processamento de Linguagem Natural, por outro lado, preocupa-se diretamente com o estudo da linguagem voltado para a construção de *softwares*, aplicativos e sistemas computacionais específicos, como tradutores automáticos. Cabe à área de PLN a construção de programas capazes de interpretar e/ou gerar informações em linguagem natural.

Uns dos desafios do Processamento de Linguagem Natural se deve à complexidade e riqueza das línguas humanas, por este motivo os estudos se concentram, cada vez mais, no desenvolvimento de tradutores automáticos.

3.4. Tradução Automática

Tradução automática (TA) é o processo automático de tradução de um idioma original para outro através do computador.

A tradução automática é uma subseção da linguística computacional que investiga o uso de programas de computador para traduzir textos ou discursos de uma língua natural a outra. No seu nível básico, a TA executa a substituição simples de palavras de uma língua natural por palavras em outra. (TRADUÇÃO AUTOMÁTICA, 2013)

A ideia de TA surgiu no século XVII quando filósofos como Descartes e Leibniz propuseram a criação de uma língua universal. Em 1954, com a experiência de Georgetown, realizou-se a tradução completa de mais de 60 frases do russo para o inglês. A experiência teve resultados positivos e iniciou uma nova era onde se investiu fortemente na pesquisa em tradução automática. Os autores consideravam que de três a cinco anos o problema de TA seria resolvido. Porém, percebeu-se que o trabalho não era trivial e após dez anos de estudos sem resultados satisfatórios, culminou em um declínio nos investimentos.

A tradução automática é uma tarefa árdua devido, principalmente, à necessidade de um conhecimento linguístico aprofundado das duas (ou mais)

línguas envolvidas para a construção de recursos, como gramáticas de tradução, dicionários bilíngues etc. (CASELI, 2007).

O processo de tradução não é constituído por uma mera substituição de cada palavra, mas "conhecer" todas as palavras de uma oração e identificar como cada uma pode interferir no significado da construção.

O grande problema das linguagens naturais é a presença de ambiguidade, até mesmo em pequenas frases. Pois, sabe-se que as línguas humanas são constituídas de morfologia, ou seja, o modo como as palavras são montadas a partir de pequenas unidades de sentido; sintaxe ou estrutura da língua; e semântica.

Os programas atuais de tradução automática muitas vezes permitem uma personalização por domínio ou profissão — melhorando os resultados por limitar o escopo de substituições admissíveis. Esta técnica é especialmente eficaz em domínios onde a língua formal é utilizada. Logo a tradução automática de documentos do governo e legais mais prontamente produz resultados mais usáveis que a conversação ou textos menos padronizados (TRADUÇÃO AUTOMÁTICA, 2013).

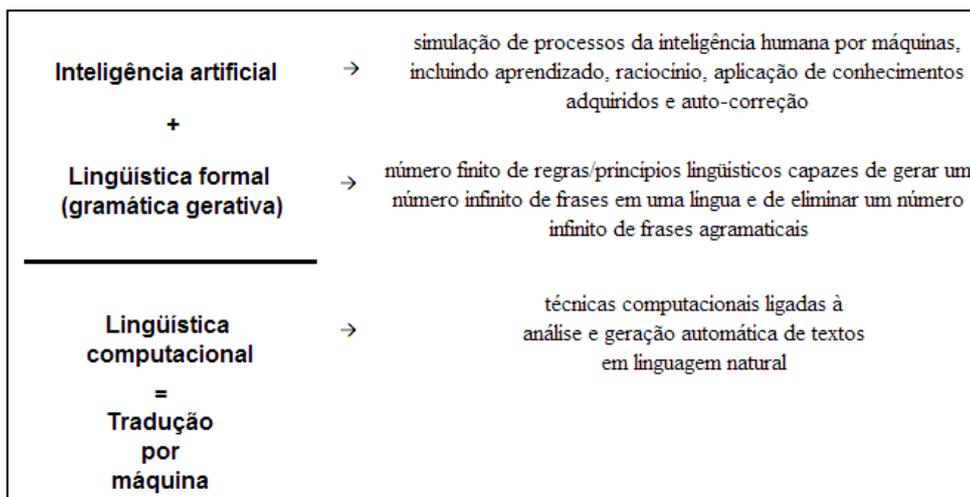
Com o objetivo de usar computadores digitais no mecanismo de tradução de línguas naturais, em 1946, Andrew Donald Booth propôs o uso de Máquinas de Tradução Automática (MTA).

3.4.1. Máquina de Tradução Automática

Apesar dos desenvolvimentos alcançados nas décadas de 50 e 60, os avanços na área das traduções automáticas ainda estavam muito aquém do desejado, assim em meados dos anos 80 surgiram vários fatores que revitalizaram o interesse pela tradução feita por máquina.

A tradução por máquina se enquadra dentro da área da linguística computacional, que pode ser definida como uma subárea da inteligência artificial combinada com a aplicação da linguística formal (em especial a gramática gerativa de Chomsky) (ALFARO & DIAS, [199-?]). Assim, segue abaixo a Figura 6.

Figura 6 - A tradução por máquina no panorama atual.



Fonte: (ALFARO & DIAS, [199-?])

Hoje existem pesquisas focadas em Máquinas de Tradução Automática (MTA), porém ainda há grandes dificuldades para um tratamento prático de casos, pois as tecnologias mais avançadas empregadas atualmente são proprietárias e utilizadas em sistemas vendidos comercialmente. São poucas e recentes as alternativas de sucesso que visam criar tecnologias abertas para pesquisa (MIRANDA JR., 2009).

Uma das MTAs de código aberto e que facilitam o estudo e o desenvolvimento de novas técnicas de tradução automática é o Apertium.

3.5. Apertium

O "Opentrad Apertium" é definido como uma plataforma de tradução automática de código aberto, baseada em regras sintáticas superficiais, com capacidade de análise morfológica, sintática e realização de transferências gramaticais por meio de regras de tradução (RAMÍREZ-SÁNCHEZ, SÁNCHEZ-MARTÍNEZ, ORTIZ-ROJAS, PÉREZ-ORTIZ, & FORCADA, 2006).

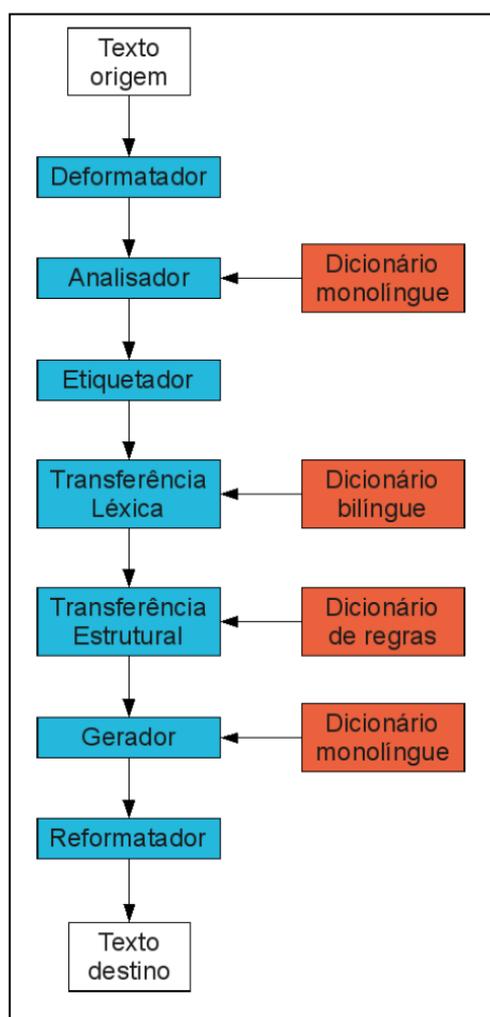
O projeto Opentrad, criado em 2004/2005, foi financiado pelo Ministério de Indústria, Turismo e Comércio da Espanha que inicialmente teve como objetivo fazer a tradução entre os dialetos espanhóis, por exemplo espanhol - catalão. Mais recentemente, o Apertium recebeu também financiamento da Generalitat de

Catalunya para desenvolver novos pares de línguas e uma máquina melhorada, Apertium 2.0, para incluir pares mais distintos como o par inglês - catalão.

O Apertium proporciona inteligência de tradução entre as línguas; ferramentas para gerenciar os dados linguísticos necessários para a construção de um sistema de tradução automática para um determinado par de línguas; e dados linguísticos para um número crescente de pares de línguas.

A MTA do Apertium é composta, basicamente, por módulos independentes e coordenados que se comunicam por meio de artefatos em formato texto (MIRANDA JR., 2009). Na Figura 7 têm-se a arquitetura simplificada do Apertium.

Figura 7 - Arquitetura simplificada do Apertium.



Fonte: (BECK, 2009)

Nos tópicos a seguir, encontra-se a função de cada módulo:

- **Deformatador:** ele é o responsável por separar o texto a ser traduzido do formato original em que o documento se encontra (HTML, RTF, etc.). As informações da formatação são encapsuladas para serem tradas posteriormente por outros módulos. Um exemplo de seria as *tags* HTML <h1> que representam títulos.
- **Analisador:** é o módulo que segmenta o texto morfológicamente retornando as formas lexicais. Cada uma das partes é composta por um lema, que é a palavra comumente utilizada nos dicionários, uma categoria morfológica, pode ser um substantivo, verbo, adjetivo, etc., e informações sobre a flexão morfológica, como número, gênero e pessoa. Cada forma léxica é determinada de acordo com o dicionário monolíngue gerando uma sequência de etiquetas (< >) que mostram a classificação morfológica daquela palavra. Por exemplo o lema "casa" pode gerar duas formas léxicas diferentes: o substantivo feminino singular "casa" (<n><f><sg>) ou o verbo "casar" na segunda pessoa do singular do modo imperativo (<vblex><imp><p2><sg>).
- **Etiquetador:** A etiquetagem é responsável por escolher a forma léxica mais adequada ao contexto diante de uma ambiguidade morfológica. Utilizando como base o exemplo do tópico anterior, o etiquetador é o responsável por escolher a forma léxica mais adequada ("casa" ou "casar") diante do lema "casa".
- **Transferência Léxica:** A transferência léxica é responsável pela tradução. Cada palavra na língua fonte é traduzida para a língua alvo de acordo com sua forma léxica, utilizando para isso o dicionário bilíngue. Por exemplo a frase "a casa", na tradução do Português para o Inglês, resultaria na frase "the house" com suas respectivas etiquetas como na forma lexical da língua fonte.
- **Transferência Estrutural:** É a responsável por detectar padrões de palavras que precisem de um tratamento especial em virtude de divergências estruturais entre o par de línguas envolvido. Como, trocas de gênero, trocas de número, entre outros. Para realizar essa tarefa, o módulo de transferência estrutural conta com o auxílio do dicionário de regras. Por exemplo, ela realiza a verificação de concordância entre artigos e substantivos.

- Gerador: O objetivo do gerador morfológico é transformar o texto que já está na língua alvo mas que, porém, ainda possui palavras nas suas respectivas formas lexicais (etiquetas), nos seus respectivos lemas. Para isso o gerador morfológico utiliza o dicionário monolíngue da língua alvo.
- Reformatador: A última etapa é gerar o texto traduzido na língua alvo para sua formatação inicial original. A formatação identificada no Deformatador. Por exemplo, aqui retornariam as *tags* HTML <h1> dos títulos.

É possível usar Apertium para construir sistemas de tradução automática para uma variedade de pares de idiomas. Para isso, o Apertium utiliza formatos simples baseados em XML para codificar os dados linguísticos necessários que são compilados, com as ferramentas que são fornecidos nos formatos de alta velocidade usada pela máquina.

4. METODOLOGIA

O desenvolvimento do presente projeto envolve um estudo aprofundado das tecnologias, bem como de toda base teórica de regras de tradução e linguística computacional. Optou-se então, por uma metodologia cíclica de pesquisa onde cada ciclo passa por um conjunto de etapas a fim de contemplar os objetivos propostos.

Pelo fato do contexto do projeto ser amplo e para alcançar os objetivos almejados listou-se uma série de etapas a fim de compor os ciclos. Abaixo, encontram-se algumas delas:

- Estudo de linguística. Como área de estudo da linguagem, a linguística cumpre importante papel já que difunde conhecimentos e descobertas da linguagem que permitem ao homem entender melhor as mudanças provenientes do seu processo de evolução dentro da história da humanidade. O estudo da linguística, principalmente da linguística computacional, é imprescindível para a realização do presente projeto. Dado que a criação de regras de tradução está diretamente ligada com a linguística computacional atual.
- Estudo de LIBRAS. O conhecimento da linguagem de sinais, hoje em dia, é de fundamental importância para o estabelecimento da comunicação entre surdos e ouvintes. Saber LIBRAS é abrir as portas para a inclusão dos surdos na sociedade brasileira atual. O entendimento de LIBRAS é importante para o projeto uma vez que é necessário noções do comportamento gramatical da linguagem para a criação das regras de tradução Português - LIBRAS. Essas regras irão proporcionar mecanismos para facilitar a difusão do conhecimento para as pessoas portadoras de algum tipo de deficiência auditiva.
- Entendimento do sistema Apertium. Tendo em vista que o sistema Apertium, é o mais adequado para o presente projeto, o estudo e o conhecimento do seu modo de funcionamento bem como das ferramentas envolvidas traz a base necessária para a implementação das regras de tradução criadas. O Apertium é uma plataforma de máquina de tradução de código aberto o que facilita seu estudo por não demandar custos com a compra de licenças, etc. É necessário, também, o entendimento de sua arquitetura para a criação do par de tradução Português - LIBRAS.

- Experimentos com o Apertium. Uma vez conhecida sua arquitetura e funcionamento, é chegada a hora de fazer experimentos com o Apertium. A fase de experimentos com o Apertium foi elaborada para que, com o contato direto com a ferramenta, fosse possível a descoberta de problemas, bem como pontos de melhoria para o projeto.
- Estudo de regras de tradução. A construção de um tradutor automático requer uma grande quantidade de conhecimento de tradução - geralmente armazenado em dicionário bilíngues bases de exemplos ou modelos estatísticos (MENEZES & RICHARDSON, 2001) - de difícil construção ou manutenção. Para tal, tem-se que o estudo de regras de tradução é essencial para o desenvolvimento do projeto.
- Recorte de cenários para teste e experimento. A importância do recorte de cenários se deve ao fato de que o entorno do projeto é bastante amplo. Delimitando um cenário para o teste e o experimento, neste caso, têm-se um resultado de maior qualidade e precisão.
- Execução de experimentos. A execução dos experimentos serve para comprovar ou refutar as hipóteses e os métodos utilizados. É importante validar os conhecimentos empregados e o resultado do projeto.
- Aplicação dos conceitos de LIBRAS para a tradução. A aplicação dos conceitos de LIBRAS na tradução facilita a criação do par de tradução e a adaptação da linguagem. Por serem duas línguas distintas, o Português e a LIBRAS, é de grande importância delimitar e estabelecer corretamente o conjunto de regras necessários para realizar a tradução de forma correta.
- Análise do resultado final. A análise do resultado final decide se o projeto alcançou os objetivos. É importante analisar se os resultados obtidos foram proveitosos, satisfatórios e se atenderam as expectativas iniciais. É importante também, para os possíveis trabalhos futuros.

A execução das etapas descritas acima serviram como meio para o entendimento e a resolução do problema. O uso de uma metodologia eficaz melhora a qualidade e o desenvolvimento do trabalho.

5. RESULTADOS

O fato de as linguagens serem vivas, extremamente dinâmicas e ambíguas é o determinante do grande desafio de fazer a tradução automática de maneira satisfatória. Ainda hoje é extremamente complexo formalizar conceitos, formas de expressão, representações e as relações entre duas línguas quaisquer (MIRANDA JR., 2009).

Desenvolver um novo par de tradução Português - LIBRAS no Apertium é uma tarefa que exige conhecimento das línguas envolvidas, bem como de linguística e da linguística computacional, além de conhecimentos específicos de computação como o uso da ferramenta Apertium e as linguagens de desenvolvimento envolvidas para a criação das regras, no caso deste trabalho, a linguagem XML.

A linguística computacional, conforme descrito da seção 3.3, é a área de conhecimento que explora as relações entre linguística e informática, tornando possível a construção de sistemas com capacidade de reconhecer e produzir informação apresentada em linguagem natural (VIEIRA & LIMA, 2001).

Uma das áreas mais estudadas da linguística computacional, é a área de processamento da linguagem natural. Na seção 3.1 tem-se que as linguagens naturais são as linguagens utilizadas como um meio de comunicação cotidiana entre os seres humanos. Visto isso, existe um grande desafio para ciência da computação que é o desenvolvimento de computadores capazes de entender a linguagem natural. O processamento da linguagem natural lida diretamente com isso, a interação entre computadores e idiomas humanos, ou linguagens naturais.

Neste sentido, as línguas gestuais também são consideradas linguagens naturais. No Brasil, conforme visto na seção 3.2, tem-se a LIBRAS (Linguagem Brasileira de Sinais) como a linguagem oficial das comunidades surdas. A LIBRAS, em especial, possui certas particularidades que dificultam o processamento da linguagem. Como por exemplo, se tratando do processo de tradução automática do Português para a LIBRAS, a dificuldade de reconhecimento das categorias das palavras.

Para se fazer a análise da estrutura das sentenças é necessário que primeiramente se faça o reconhecimento das categorias. Os sistemas que realizam este tipo de tarefa são denominados etiquetadores de categorias gramaticais: dado

um texto, esse texto é devolvido com o acréscimo, a cada palavra, de uma etiqueta com informação a respeito de sua categoria gramatical.

Uma vez reconhecida a categoria de uma palavra, o próximo passo na análise da linguagem natural, é verificar se a estrutura das frases é válida e reconhecer, dentro dessa estrutura, os constituintes da frase.

O sistema escolhido para realizar a tradução do Português - LIBRAS foi o sistema Apertium. O Apertium é uma plataforma livre/de código-fonte aberto para a realização de tradução automática. Na seção 3.5 tem-se especificado seu modo de funcionamento básico bem como sua arquitetura interna.

O objetivo deste capítulo é, então, exibir e formalizar a sequência de passos realizados para a especificação do conhecimento associado à construção do par de línguas Português - LIBRAS no Apertium.

Para tanto, é importante introduzir alguns conceitos necessários para a compreensão dos resultados do trabalho e da estrutura básica dos arquivos criados.

5.1. Conceitos básicos relacionados ao Apertium

Existem vários conceitos envolvidos na compreensão do sistema e arquitetura Apertium. Nesse primeiro momento, é necessário destacar dois conceitos julgados mais importantes: Lexema e Símbolo. Lexema é um dos conceitos de linguística e consiste em uma palavra ou parte de uma palavra que não possui qualquer informação gramatical. Em Português, pode-se citar como exemplos de lexemas, os lexemas de substantivos que são tipicamente representados na sua forma singular e no masculino e os lexemas de verbos que são representados em sua forma infinitiva (MIRANDA JR., 2009).

O outro conceito importante a ser destacado, é o conceito de “Símbolo”, que no contexto Apertium, é empregado em etiquetas de marcação ou de classificação. As etiquetas de marcação ou de classificação podem ser gramaticais ou podem servir apenas como auxílio às etapas de execução da máquina. A classificação das palavras, no Apertium, nada mais é do que uma sequência de etiquetas associadas. Nas entradas e saídas dos módulos do Apertium estas etiquetas são tipicamente delimitadas por “< >” e representadas por abreviações. Em resumo, os símbolos são

utilizados para fornecer características morfológicas e gramaticais. Nem todas as línguas possuem as mesmas características gramaticais, por este motivo a lista de símbolos utilizados variam de língua para língua. No Quadro 1, apresenta-se como exemplo alguns dos símbolos criados/utilizados na LIBRAS.

Quadro 1 - Exemplos de Símbolos utilizados no Apertium

Símbolo	Definição
<n>	Substantivo
<adj>	Adjetivo
<prn>	Pronome
<p1>	Primeira Pessoa
<p2>	Segunda Pessoa
<p3>	Terceira pessoa
<inf>	Infinitivo
<past>	Passado
<fut>	Futuro

Fonte: o autor.

O conceito de Paradigma também é uns dos conceitos relevantes que precisam ser especificados. Paradigma se refere a um roteiro de flexões morfológicas para um grupo de palavras. Dentro dos dicionários monolíngues, os lexemas são em sua maioria ligados diretamente a paradigmas, para que todas as variações do lexema sejam descritos sem necessidade de reescrever todas as formas de flexão (MIRANDA JR., 2009).

Exemplificando, um dos paradigmas presentes no dicionário de Português é relacionado aos verbos regulares. Como se sabe, os verbos regulares são aqueles verbos que flexionam da mesma maneira. Assim no Apertium criou-se apenas um paradigma que representa a flexão desses verbos e todos os verbos regulares foram associados a este paradigma.

5.2. Desenvolvendo um novo par de línguas no Apertium

Como o Apertium é uma máquina de tradução genérica, podem ser criados novos pares de línguas por meio do desenvolvimento de dicionários e regras de transferências. Conforme visto na seção 3.5, uma vez desenvolvidos esses

componentes é possível realizar a tradução de textos operando sobre suas unidades léxicas.

Além do motor em si, cada par de línguas do Apertium é composto por quatro dicionários, isto é, cinco arquivos em formato XML que reúnem as informações linguísticas necessárias daquele par (BECK, 2009).

Portanto, para a realização deste trabalho, foram criados cinco arquivos para o desenvolvimento do par Português - LIBRAS, como ilustrado no Quadro 2. Esses cinco arquivos funcionam como "combustível" alimentando a máquina de tradução.

Outros arquivos também podem ser utilizados para aprimorar o processo de tradução, porém eles fogem do escopo deste trabalho.

Quadro 2 - Arquivos utilizados na tradução Português-LIBRAS

Tipo	Conteúdo	Línguas	Nome
Dicionários Monolíngues	Possui os lexemas e paradigmas da própria língua.	Português	apertium-pt_br-libras.pt_br.dix
		LIBRAS	apertium-pt_br-libras.libras.dix
Dicionário Bilíngue	Responsável pela correspondência entre os lexemas e símbolos das duas línguas.	Bidirecional Português - LIBRAS	apertium-pt_br-libras.pt_br-libras.dix
Regras de Transferência	Define como os lexemas são ordenados, realizam a concordância de gênero, número e grau, inserem e eliminam símbolos.	Português	apertium-pt_br-libras.pt_br-libras.t1x
		LIBRAS	apertium-pt_br-libras.libras-pt_br.t1x

Fonte: o autor.

São necessários dois dicionários monolíngues, um para cada língua do par. Optou-se por utilizar, para a língua portuguesa, o dicionário monolíngue desenvolvido pela comunidade Apertium do par de língua Português - Espanhol uma vez que, ele foi julgado mais completo e mais adequado para utilização neste trabalho.

O dicionário monolíngue da LIBRAS, teve que ser criado. Já que, até o momento não se tinha nenhum par de tradução relacionando a LIBRAS no sistema Apertium.

Os dicionários monolíngues são utilizados nas etapas de análise e geração e, são eles que possuem as informações morfológicas das palavras bem como os paradigmas de flexão. Eles são responsáveis por identificar e classificar morfológicamente as frases, expressões e palavras de uma língua.

Os dicionários bilíngues fazem as correspondências entre lexemas e símbolos das duas línguas envolvidas. O seu conteúdo mantém o conhecimento para as duas vias de tradução. Seu principal conteúdo é a relação de correspondência “palavra a palavra” entre as línguas em questão (MIRANDA JR., 2009).

Os arquivos que contêm as regras de tradução são utilizados na transferência estrutural e existe um para cada direção que se deseja traduzir. Eles nada mais são do que um conjunto de regras que definem as eventuais modificações nas estruturas de cada sentença, por exemplo a ordem em que as palavras aparecem. Essas regras podem ser muito complexas já que suportam o uso de expressões regulares e instruções condicionais como *ifs*, *for* e etc. São essas regras que fazem a correspondência e os ajustes entre as duas línguas.

5.3. Dicionários Monolíngues

Como visto anteriormente, os dicionários monolíngues são os responsáveis pela informações morfológicas das palavras através dos símbolos e paradigmas. Nesta seção será trabalhada a especificação dos arquivos criados bem como a gramática da língua envolvida e os conceitos ligados a linguagem XML. O foco aqui, será o dicionário monolíngue de LIBRAS uma vez que, visto na seção 5.2, o dicionário monolíngue de Português utilizado foi o dicionário do par Português - Espanhol criado pela comunidade Apertium e não cabe a este trabalho analisar o modo como este foi planejado.

Os dicionários monolíngues em geral, iniciam com o elemento <dictionary>, conforme demonstrado no Código 1, e é este elemento que delimita todo conteúdo

do dicionário. O conteúdo dos dicionários podem ser divididos em quatro partes: alfabeto, símbolos, paradigmas e seções.

Código 1 - Definição base de dicionários

```
<?xml version="1.0" encoding="UTF-8"?>
<dictionary>
    //Conteúdo do dicionário
</dictionary>
```

Fonte: (MIRANDA JR., 2009)

A primeira etapa é definir o alfabeto, ou seja, os caracteres válidos para a língua. A definição do alfabeto é feito através do elemento <alphabet>. O alfabeto da LIBRAS é praticamente o mesmo do alfabeto português, porém ele possui um caractere a mais que é o caractere "@". Conforme visto na seção 3.2, na LIBRAS não há flexão de gênero, ou seja, masculino e feminino, e o sinal que representa estas marcas na língua portuguesa é o símbolo "@" para reforçar a ideia de ausência e não haver confusão para o leitor.

Porém, durante o desenvolvimento, deparou-se com a impossibilidade de utilizar no Apertium o caractere "@" como parte do alfabeto da LIBRAS. Isso se deve ao fato do "@" ser um caractere reservado no sistema Apertium. Optou-se, então, pela substituição do "@" pelo "ñ" já que este é um símbolo que também não pertence ao alfabeto português.

No Código 2 tem-se o alfabeto definido para o dicionário monolíngue de LIBRAS.

Código 2 - Alfabeto do dicionário monolíngue de LIBRAS

```
<alphabet>
ÀÁÂÃÄÅÇÈÉÊËÌÍÎÏÐÒÓÔÕÖÙÚÛÜàáâãäåçèéêëìíîïðóôõöùúûüABCDEF GHIJ
KLMNOPQRSTUVWXYZabc defghijklmnopqrst uvwxyzñ
</alphabet>
```

Fonte: o autor.

Na sequência, a segunda etapa é a classificação dos símbolos. A classificação é delimitada pelo elemento <sdefs>. O Código 3 é um trecho de código extraído do dicionário monolíngue de LIBRAS.

Código 3 - Exemplo de classificação dos símbolos para a LIBRAS

```
<sdefs>
  <sdef n="adj" c="adjetivo"/>
  <sdef n="prn" c="pronome"/>
  <sdef n="pres" c="presente"/>
  <sdef n="past" c="passado"/>
  <sdef n="fut" c="futuro"/>
  ...
  <sdef n="p1" c="1ª pessoa"/>
</sdefs>
```

Fonte: o autor.

A terceira etapa é a definição dos paradigmas. Os paradigmas servem para agrupar as palavras que flexionam do mesmo modo. O Código 4 possui o exemplo de um paradigma desenvolvido no dicionário de LIBRAS para as palavras EL@ (elñ) e EL@S (elñs). Os paradigmas devem estar contidos dentro do elemento <pardefs>. É importante ressaltar que a ordem com que as informações morfológicas aparecem no paradigma corresponde à ordem em que as etiquetas serão colocadas na forma léxica da palavra.

Código 4 - Paradigma para as palavras EL@ e EL@s

```

<pardef n="el/ñ__prn">
  <e>
    <p>
      <l/>
      <r><s n="prn"/><s n="p3"/><s n="sg"/></r>
    </p>
  </e>
  <e>
    <p>
      <l>s</l>
      <r><s n="prn"/><s n="p3"/><s n="pl"/></r>
    </p>
  </e>
</pardef>

```

Fonte: o autor.

O próximo elemento é o elemento <section>. Ele constitui a seção de lexemas da língua. Um atributo importante de <section> é "type", que define o tipo de conteúdo da seção e pode receber dois valores: "standard", simbolizando que o conteúdo é tipicamente de lexemas e "incondicional" simbolizando a definição de pontuações e outros símbolos (MIRANDA JR., 2009).

No caso do dicionário de LIBRAS foi utilizado o <section> de valor "standard" conforme o Código 5.

Código 5 - Exemplo <section> dicionário LIBRAS

```

<section id="main" type="standard">
  <e lm="elñ">
    <i>elñ</i>
    <par n="el/ñ__prn"/>
  </e>
</section>

```

Fonte: o autor.

O dicionários monolíngues de LIBRAS, utilizados neste trabalho, em sua íntegra encontra-se no Apêndice A.

5.4. Dicionário Bilíngue

O dicionário bilíngue é utilizado na transferência léxica. A estrutura do dicionário bilíngue é a mesma dos dicionários monolíngues e o elemento <sdefs> deve conter todos os símbolos das duas línguas envolvidas.

É o <section> que contém cada o lema na língua fonte (Português) e seu correspondente na língua destino (LIBRAS), conforme mostra o Código 6. Sua função é gerenciar a troca de padrões e não havendo a presença de ambiguidades, ele funciona como um tradutor "palavra a palavra".

Código 6 - Exemplo do mapeamento de palavras Português - LIBRAS

```
<section id="main" type="standard">
  <e><p>
    <l>ele<s n="prn"/></l>
    <r>elñ<s n="prn"/></r>
  </p></e>
</section>
```

Fonte: o autor.

Todos os dicionários são compilados previamente. A ideia da compilação prévia é melhorar o desempenho do acesso ao dicionário. O dicionário bilíngue, desenvolvido neste trabalho, pode ser visto em sua íntegra no Apêndice B.

5.5. Regras de Tradução

As regras de tradução são utilizadas na transferência estrutural e existe um arquivo para cada direção que se deseja traduzir. Por exemplo, existe um arquivo para a tradução do Português para LIBRAS e outro arquivo da LIBRAS para o Português. O escopo desse trabalho engloba a tradução do Português para a

LIBRAS, deixando assim, a criação de regras de tradução da LIBRAS para o Português como proposta de um trabalho futuro.

Esse arquivo contendo as regras de tradução do Português para LIBRAS é composto por um conjunto de regras que definem eventuais modificações na estrutura das sentenças. Um exemplo da estrutura básica desse arquivo é dado no Código 7.

Código 7 - Estrutura básica do arquivo de Regras de Transferência

```
<?xml version="1.0" encoding="UTF-8"?>
  <transfer>
    <section-def-cats>
      <!-- Definição de Categorias -->
    </section-def-cats>
    <section-def-attrs>
      <!-- Definição de Atributos -->
    </section-def-attrs>
    <section-def-vars>
      <!-- Definição de Variáveis -->
    </section-def-vars>
    <section-rules>
      <!-- Definição de Regras -->
    </section-def-rules>
  </transfer>
```

Fonte: (MIRANDA JR., 2009)

Inicialmente, havia sido pensado para este trabalho o recorte de um cenário para realizar a tradução. O cenário escolhido seria uma pequena historia infantil. Porém, devido ao tempo disponível, não foi possível criação das regras de tradução necessárias. Optou-se então por traduzir frases curtas mas suficientemente complexas para servir como resultado satisfatório deste projeto.

Para fins de exemplificação e demonstração do trabalho realizado, a frase em Português escolhida foi "Corri.". A escolha da frase pode ser justificada por vários

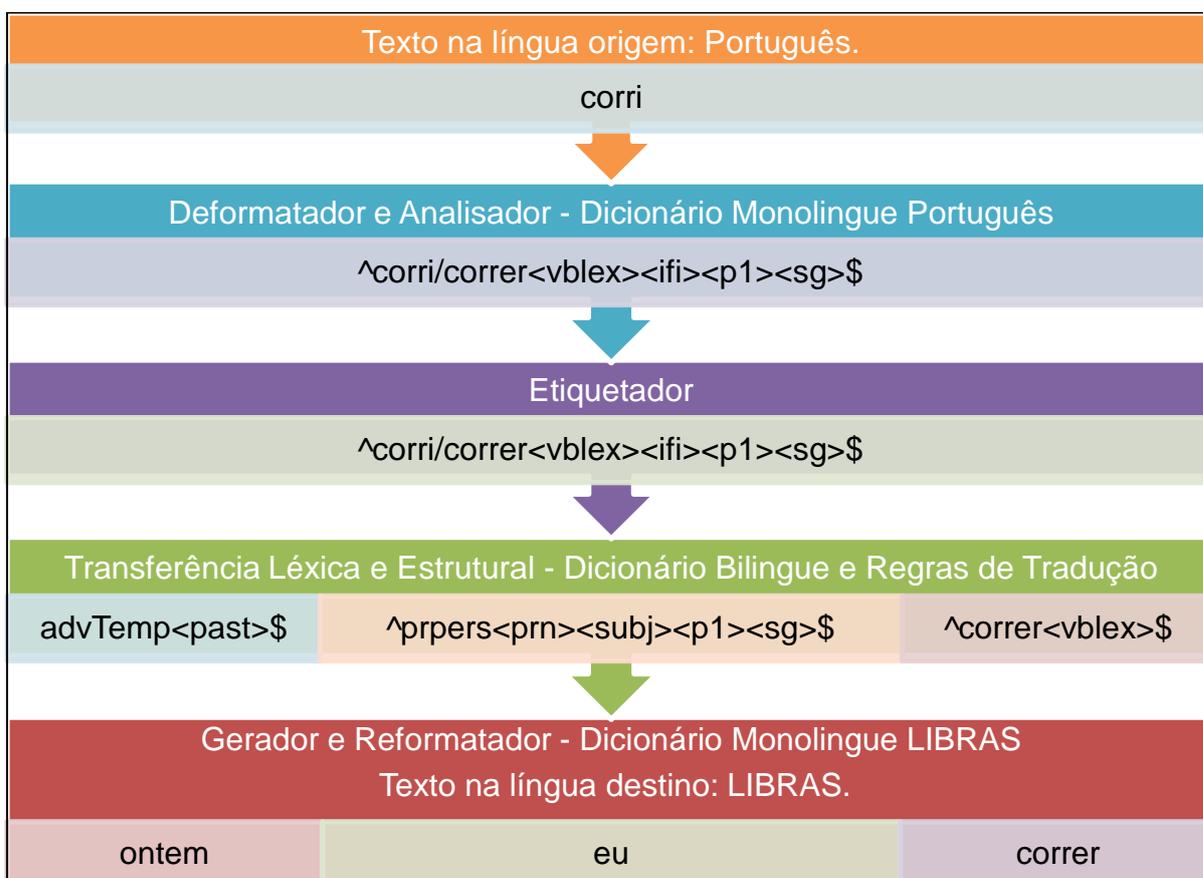
motivos gramaticais, visto na seção 3.2, dentre eles podemos citar o fato de que na LIBRAS:

- Não há marca de tempo nas formas verbais, nas frases, muitos verbos ficam no infinitivo.
- O pronome pessoal deve ser explícito para representar as pessoas do discurso.

Em Português quando uma pessoa fala "Corri" é possível compreender, imediatamente, que essa pessoa locutora correu em algum momento passado. Porém na LIBRAS não é tão trivial. Assim a frase "Corri" traduzida do Português para a LIBRAS seria: "Ontem eu correr."

Na seção a seguir encontra-se as regras de tradução desenvolvidas para obter o resultado esperado. Essas regras seguem um processo conforme ilustrado na Figura 8.

Figura 8 - Processo de Tradução da Frase



Fonte: o autor.

5.5.1. Criando Regras de Tradução no Apertium

Para criar as regras de tradução, primeiramente é necessário definir as categorias e os atributos que permitem que os símbolos gramaticais sejam agrupados. As categorias permitem agrupar os símbolos com mesmo padrão, como em "vblex.*", que é o padrão para todos os verbos. Os atributos contêm os símbolos disponíveis para uma determinada opção. Por exemplo, os símbolos "passado" (past) e "futuro" (fut) podem ser agrupados em um atributo "temps". O Código 8 demonstra a categoria e o atributo criado para esses exemplos.

Código 8 - Exemplo definição de Categoria e Atributo

```
<def-cat n="vrb">
  <cat-item tags="vblex.*"/>
</def-cat>
<def-attr n="temps">
  <attr-item tags="fut"/>
  <attr-item tags="past"/>
</def-attr>
```

Fonte: o autor.

Uma vez definidos todas as categorias e atributos já é possível desenvolver as regras de tradução. A regra criada diz respeito à tradução da frase em português "Corri." para a frase em LIBRAS "Ontem eu correr.". Essa regra, pode ser dividida em três partes, para facilitar o entendimento.

A primeira parte diz respeito à substituição dos tempos verbais, ou a troca dos símbolos temporais da frase. "Corri." em português é etiquetado como <ifi> que significa o tempo verbal Pretérito Perfeito do Indicativo. Para a frase na LIBRAS, esse símbolo deve ser substituído por <past> que indica, simplesmente, o tempo Passado, já que na LIBRAS não se tem diferenciação de tipos de passado. O Código 9 é um trecho extraído da regra responsável por essa modificação.

Código 9 - Substituição de Símbolos

```
<let>
  <clip pos="1" side="tl" part="tense"/>
  <lit-tag v="past"/>
</let>
```

Fonte: o autor.

A segunda parte se refere à inserção do advérbio de tempo na frase. Os advérbios de tempo na LIBRAS geralmente vêm no começo da frase. Um dos advérbios de tempo responsáveis por indicar uma ação que ocorreu no passado é o ONTEM. O Código 10 é o responsável por essa inserção do advérbio de tempo no começo da frase.

Código 10 - Inserção do Advérbio de Tempo

```
<lu>
  <lit v="advTemp"/>
  <clip pos="1" side="tl" part="temps"/>
</lu>
```

Fonte: o autor.

A terceira parte trata da inclusão do pronome pessoal e a modificação do verbo para sua forma no infinitivo. O que ocorre é que ao encontrar um verbo em Português, a frase é reordenada e é criado um novo lexema "prpers" auxiliado pelos atributos de pessoa e número que estavam presentes na flexão do verbo e o verbo volta à sua forma infinitiva. No Código 11 é demonstrado como essa modificação é feita.

Código 11 - Inclusão do Pronome Pessoal

```

<lu>
  <lit v="prpers"/>
  <lit-tag v="prn"/>
  <lit-tag v="subj"/>
  <clip pos="1" side="t1" part="pers"/>
  <clip pos="1" side="t1" part="nbr"/>
</lu>
<b/>
<lu>
  <clip pos="1" side="t1" part="lem"/>
  <clip pos="1" side="t1" part="a_verb"/>
</lu>

```

Fonte: o autor.

O arquivo de regras de tradução do Português para a LIBRAS, desenvolvido neste trabalho, pode ser conferido na íntegra no Apêndice C.

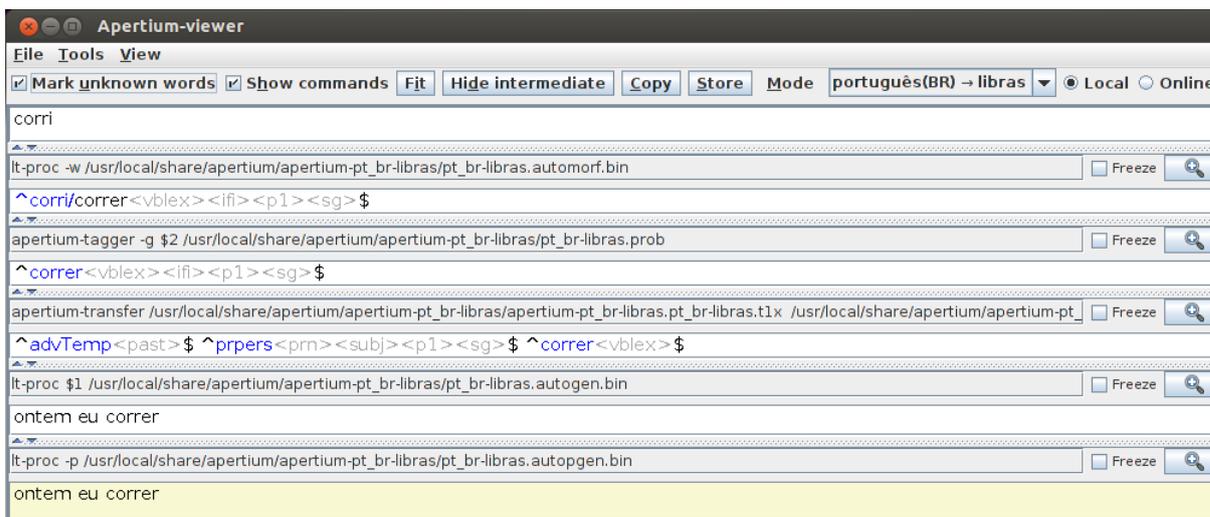
5.6. Considerações Finais

A ideia de descrever o processo é apresentar uma parcela dos desafios da tarefa de especificar um novo par de línguas no Apertium, principalmente se tratando de línguas tão diferentes, uma da outra, como o Português e a LIBRAS.

Vale ressaltar que o processo pode ser muito mais complicado, compreendendo outras fases, estruturas e arquivos. Porém o resultado que foi obtido com este trabalho atingiu o objetivo proposto, uma vez que, com esses cinco arquivos já é possível efetuar os primeiros passos na tradução de sentenças. A Figura 9 mostra o funcionamento da tradução da frase escolhida utilizando o Apertium-viewer.

O Apertium-viewer é um pequeno programa que pode ser usado para visualizar e editar a saída das várias fases de uma tradução Apertium. As várias etapas são atualizadas enquanto você digita, e uma alteração feita no painel de qualquer uma das fases, atualiza as fases subsequentes.

Figura 9 - Tradução no Apertium-viewer



Fonte: o autor.

Este trabalho pode ser considerado um grande passo rumo à tradução por completa da Língua Portuguesa para a LIBRAS.

6. CONCLUSÕES E PROPOSTAS DE CONTINUIDADE

A LIBRAS é uma língua completa. Ou seja, um surdo pode expressar-se verdadeiramente pela língua de sinais se souber essa língua tanto quanto a um ouvinte que domine sua língua oral. É necessário então, criar mecanismos que facilitem a sua compreensão e entendimento.

Um dos maiores problemas relacionados com a inclusão social do surdo na sociedade brasileira é o estabelecimento da comunicação entre os surdos e os ouvintes. Vê-se então, que através da linguística computacional juntamente com as máquinas de tradução automática será possível desenvolver regras de tradução para a criação de um tradutor Português - LIBRAS. Esse tradutor incentivaria não só o conhecimento do Português por parte dos surdos, mas também da LIBRAS por parte dos ouvintes.

A utilização da metodologia cíclica com as primeiras etapas voltadas para o estudo e a preparação do ambiente de desenvolvimento é satisfatória uma vez que fornece o embasamento teórico necessário para o desenvolvimento do projeto. Através do estudo de linguagens natural, máquinas de tradução automática, linguística computacional, bem como da LIBRAS é possível a formulação de teorias e hipóteses.

A escolha do Apertium como ferramenta deve-se ao fato de ser uma plataforma livre e que se encaixa nas ideias propostas. O Apertium é uma ferramenta importante e forte na área de tradutores automáticos. Mesmo com alguns pontos negativos, por exemplo a ausência de uma interface amigável, através do estudo sistemático de sua arquitetura é possível desenvolver trabalhos que contribuem para sua evolução.

O tradutor Português - LIBRAS pretende ser a ponte que liga o mundo fechado e individual do surdo com o mundo de conhecimento e da ciência que muitas vezes não está acessível a ele devido à sua condição de deficiência.

Como propostas de continuidade do presente projeto pode-se citar:

- A criação do par inverso de tradução com o objetivo de realizar a tradução tanto do Português para LIBRAS como da LIBRAS para Português. Assim, há uma maior contribuição para o estabelecimento do diálogo e a passagem do conhecimento.

- Ampliação das regras de tradução com o objetivo de aumentar a abrangência da tradução. Na fase inicial de um tradutor, é escolhido cenários para que a tradução seja realizada criando assim, regras específicas daquele cenário. Com a ampliação das regras de tradução, será possível envolver outros cenários, até mesmo, contemplar a língua por inteiro.
- Criação de ambiente mais desenvolvido com o objetivo de facilitar o uso por parte de usuários leigos. Possivelmente desenvolver uma interface amigável e com alta usabilidade a fim de ampliar o público alvo da ferramenta.

REFERÊNCIAS

- ALFARO, C., & DIAS, M. ([199-?]). *SISTEMAS DE TRADUÇÃO POR MÁQUINA: FERRAMENTAS DE AUXÍLIO À TRADUÇÃO*. Tecgraf PUC Rio. Disponível em: <<http://www.tecgraf.puc-rio.br/~carolina/ferramentas.html>>. Acesso em: 8 de Agosto de 2013.
- BECK, D. (Dezembro de 2009). Aprimorando o tratamento de Expressões Multipalavras em um tradutor automático baseado em regras. Porto Alegre.
- CASELI, H. (2007). *Indução de léxicos bilíngües e regras para a tradução automática*. São Carlos.
- CHAPLIN, C. (s.d.). *Frases de Charles Chaplin*. Frases e Provérbios. Disponível em: <<http://www.frasese proverbios.com/frases-de-charles-chaplin.php>>. Acesso em: 9 de Agosto de 2013.
- CORADINE, L., ALBURQUEQUE, F., SILVA, A., MADEIRO, J., & PEREIRA, M. (2004). *Interpretação de Orações Simples Através do Falibras (Tradutor Português-Libras) para Auxiliar na Aprendizagem de Crianças Surdas*. Manaus.
- FAQUETI, C., GRANDI, G., FANTINI, L., & LORENZETTI, M. (2005). *InfoLIBRAS – O Uso da Web para o Aprendizado da Língua de Sinais com Termos da Informática*. São Leopoldo.
- FELIPE, T. (2007a). Descrição da língua de sinais: desafios teóricos e práticos. *Congresso Internacional do INES*.
- FELIPE, T. (2007b). *Libras em Contexto*. Brasília: WalPrint Gráfica e Editora.
- FERNANDES, J. (Maio de 2002). *Linguagens Naturais*. Disponível em: <<http://www.cic.unb.br/~jhcf/MyBooks/ic/3.ProgramacaoBasica/AspectosTeoricos/MaquinasLinguagens/LinguagensNaturais/linguagensnaturais.html>>. Acesso em: 27 de Agosto de 2013.
- FRANCO, N., SILVA, W., & BRITO, P. ([201-?]). *FALIBRASWEB: Uma Proposta de Inclusão Social para Surdos*. Alagoas.

HUTCHINS, J., & SOMERS, H. (1992). *An introduction to machine translation*. Disponível em: <<http://www.hutchinsweb.me.uk/IntroMT-TOC.html>>. Acesso em: 7 de Agosto de 2013.

IBGE. (2010). *Tabelas de Resultados do Censo Demográfico*. IBGE. Disponível em: <ftp://ftp.ibge.gov.br/Censos/Censo_Demografico_2010/Caracteristicas_Gerais_Religiao_Deficiencia/tab1_3.pdf>. Acesso em: 22 de Julho de 2013.

LÍNGUA BRASILEIRA DE SINAIS. (2013). WIKIPÉDIA. Disponível em: <http://pt.wikipedia.org/wiki/L%C3%ADngua_brasileira_de_sinais>. Acesso em: 6 de Agosto de 2013.

MENEZES, A., & RICHARDSON, D. (2001). A best-first alignment algorithm for automatic extraction of transfer mappings from bilingual corpora. Toulouse, France.

MIRANDA JR., A. (2009). *WiKLaTS – UM AMBIENTE DE INTERFACE E INTERAÇÃO PARA MANIPULAÇÃO E FORMALIZAÇÃO DE CONHECIMENTO*. Curitiba: Universidade Federal do Paraná.

OTHERO, G. (Junho de 2006). *Linguística Computacional: uma breve introdução*. Porto Alegre, Rio Grande do Sul, Brasil.

PEREIRA, A. ([20--]). *Língua Brasileira de Sinais*. Brasil Escola. Disponível em: <<http://meuartigo.brasilecola.com/educacao/lingua-brasileira-sinais.html>>. Acesso em: 6 de Agosto de 2013.

RAMÍREZ-SÁNCHEZ, G., SÁNCHEZ-MARTÍNEZ, F., ORTIZ-ROJAS, S., PÉREZ-ORTIZ, J., & FORCADA, M. (2006). *Opentrad Apertium open-source machine translation system: an opportunity for business and research*. London.

TRADUÇÃO AUTOMÁTICA. (2013). WIKIPÉDIA. Disponível em: <http://pt.wikipedia.org/wiki/Tradu%C3%A7%C3%A3o_autom%C3%A1tica>. Acesso em: 8 de Agosto de 2013.

VIEIRA, R., & LIMA, V. (2001). *Linguística computacional: princípios e aplicações*. Rio Grande do Sul, Brasil.

APÊNDICE A - DICIONÁRIO MONOLÍNGUE LIBRAS

Definição completa do arquivo `apertium-pt_br-libras.libras.dix`, que define o dicionário monolíngue de LIBRAS, utilizado neste trabalho.

```
<?xml version="1.0" encoding="UTF-8"?>
<dictionary>
<alphabet>
  ÀÁÂÃÄÇÈÉÊËÏÎÏÏÑÒÓÔÕÖÙÚÛÜääåääçèéëëïíîïñòóôõöùúüABCDEF GHIJKLMNOPQRSTUVWXYZab
  cdefghijklmnopqrstuvwxyzñ
</alphabet>
<sdefs>
  <sdef n="n" c="substantivo"/>
  <sdef n="adj" c="adjetivo"/>
  <sdef n="vblex" c="verbo sem complemento (infinitivo)"/>
  <sdef n="adv" c="adverbio"/>
  <sdef n="preadv" c="pre adverbio"/>
  <sdef n="posadv" c="pos adverbio"/>
  <sdef n="prn" c="pronome"/>
  <sdef n="pr" c="preposição"/>
  <sdef n="num" c="numeral"/>
  <sdef n="sg" c="singular"/>
  <sdef n="pl" c="plural (todos)"/>
  <sdef n="du" c="dual"/>
  <sdef n="tri" c="*trial"/>
  <sdef n="quad" c="*quadrial"/>
  <sdef n="pres" c="presente"/>
  <sdef n="past" c="passado"/>
  <sdef n="fut" c="*futuro"/>
  <sdef n="aff" c="afirmativo"/>
  <sdef n="itg" c="interrogativo"/>
  <sdef n="neg" c="negativo"/>
  <sdef n="p1" c="1 pessoa"/>
  <sdef n="p2" c="2 pessoa"/>
  <sdef n="p3" c="3 pessoa"/>
  <sdef n="aa" c="animado"/>
  <sdef n="an" c="inanimado"/>
  <sdef n="SN" c="sintagma nominal"/>
  <sdef n="SA" c="sintagma adjetivo"/>
  <sdef n="SV" c="sintagma verbal"/>
  <sdef n="tn" c="tonico"/>
  <sdef n="subj" c="sujeito"/>
</sdefs>
<pardefs>
<pardef n="el/ñ__prn">
  <e>
    <p>
      <l></l>
      <r><s n="prn"/><s n="tn"/><s n="p3"/><s n="sg"/></r>
    </p>
  </e>
  <e>
    <p>
      <l>s</l>
      <r><s n="prn"/><s n="tn"/><s n="p3"/><s n="pl"/></r>
    </p>
  </e>
</pardef>
<pardef n="aquele/ñ__prn">
  <e>
    <p>
      <l></l>
      <r><s n="prn"/><s n="tn"/><s n="sg"/></r>
    </p>
  </e>

```

```

<e>
  <p>
    <l>s</l>
    <r><s n="prn"/><s n="tn"/><s n="pl"/></r>
  </p>
</e>
</pardef>
<pardef n="correr__vblex">
<e>
  <p>
    <l></l>
    <r><s n="vblex"/></r>
  </p>
</e>
</pardef>
<pardef n="prsubj__prn">
<e>
  <p>
    <l>eu</l>
    <r>prpers<s n="prn"/><s n="subj"/><s n="pl"/><s n="sg"/></r>
  </p>
</e>
</pardef>
<pardef n="advTemp_adv">
<e>
  <p>
    <l>ontem</l>
    <r>advTemp<s n="past"/></r>
  </p>
</e>
<e>
  <p>
    <l>amanhã</l>
    <r>advTemp<s n="fut"/></r>
  </p>
</e>
</pardef>
</pardefs>
<section id="main" type="standard">
  <e lm="aque!n">
    <i>aque!n</i>
    <par n="aque!n__prn"/>
  </e>
  <e lm="eln">
    <i>eln</i>
    <par n="el/n__prn"/>
  </e>
  <e lm="correr">
    <i>correr</i>
    <par n="correr__vblex"/>
  </e>
  <e lm="personal subject pronouns">
    <i></i>
    <par n="prsubj__prn"/>
  </e>
  <e lm="temporal adverbs">
    <i></i>
    <par n="advTemp_adv"/>
  </e>
</section>
</dictionary>

```

APÊNCIDE B - DICIONÁRIO BILÍNGUE PORTUGUÊS - LIBRAS

Definição completa do arquivo `apertium-pt_br-libras.pt_br-libras.dix`, que define o dicionário bilíngue Português - LIBRAS, utilizado neste trabalho.

```
<?xml version="1.0" encoding="UTF-8"?>
<dictionary>
  <alphabet></alphabet>
  <sdefs>
    <sdef n="tn" />
    <sdef n="ind" />
    <sdef n="atn" />
    <sdef n="enc" />
    <sdef n="pro" />
    <sdef n="acr" />
    <sdef n="detnt" />
    <sdef n="ant" />
    <sdef n="predet" />
    <sdef n="ND" />
    <sdef n="GD" />
    <sdef n="loc" />
    <sdef n="n" />
    <sdef n="np" />
    <sdef n="al" />
    <sdef n="adj" />
    <sdef n="f" />
    <sdef n="m" />
    <sdef n="mf" />
    <sdef n="sg" />
    <sdef n="pl" />
    <sdef n="adv" />
    <sdef n="preadv" />
    <sdef n="pr" />
    <sdef n="prn" />
    <sdef n="rel" />
    <sdef n="aa" />
    <sdef n="an" />
    <sdef n="nn" />
    <sdef n="itg" />
    <sdef n="det" />
    <sdef n="dem" />
    <sdef n="def" />
    <sdef n="cnjcoo" />
    <sdef n="cnjsub" />
    <sdef n="cnjadv" />
    <sdef n="nt" />
    <sdef n="vbser" />
    <sdef n="vbhaver" />
    <sdef n="vblex" />
    <sdef n="vbmod" />
    <sdef n="inf" />
    <sdef n="ger" />
    <sdef n="pp" />
    <sdef n="pri" />
    <sdef n="p1" />
    <sdef n="p2" />
    <sdef n="p3" />
    <sdef n="pii" />
    <sdef n="ifi" />
    <sdef n="fti" />
    <sdef n="cni" />
    <sdef n="prs" />
    <sdef n="pis" />
    <sdef n="fts" />
    <sdef n="imp" />
  </sdefs>
</dictionary>
```

```

<sdef n="infps" />
<sdef n="pmp" />
<sdef n="pos" />
<sdef n="ref" />
<sdef n="sup" />
<sdef n="num" />
<sdef n="ij" />
<sdef n="sp" />
<sdef n="sent" />
<sdef n="cm" />
<sdef n="lpar" />
<sdef n="rpar" />
<sdef n="lquest" />
<sdef n="percent" />
</sdefs>
<pardefs>
  <pardef n="numerals">
    <e>
      <p>
        <l>uno<s n="num"/></l>
        <r>um<s n="num"/></r>
      </p>
    </e>
  </pardef>
</pardefs>
<section id="main" type="standard">
  <e><p>
    <l>ele<s n="prn"/><s n="tn"/><s n="p3"/><s n="m"/><s n="sg"/></l>
    <r>elñ<s n="prn"/><s n="tn"/><s n="p3"/><s n="sg"/></r>
  </p></e>
  <e><p>
    <l>aquale<s n="prn"/></l>
    <r>aquelñ<s n="prn"/></r>
  </p></e>
  <e><p>
    <l>correr<s n="vblex"/></l>
    <r>correr<s n="vblex"/></r>
  </p></e>
</section>
<section id="final" type="standard">
  <e><re>[0-9]+([.][0-9]+)?</re><p>
    <l><s n="num"/></l>
    <r><s n="num"/></r>
  </p></e>
</section>
</dictionary>

```

APÊNDICE C - REGRAS DE TRANSFERÊNCIA PORTUGUÊS - LIBRAS

Definição completa do arquivo `apertium-pt_br-libras.pt_br-libras.t1x`, que define as regras de tradução Português - LIBRAS, utilizado neste trabalho.

```
<?xml version="1.0" encoding="UTF-8"?>
<transfer>
  <section-def-cats>
    <!-- Definição de Categorias -->
    <def-cat n="nom">
      <cat-item tags="n.*"/>
    </def-cat>
    <def-cat n="vrb">
      <cat-item tags="vblex.*"/>
    </def-cat>
    <def-cat n="prpers">
      <cat-item lemma="prpers" tags="prn.*"/>
    </def-cat>
    <def-cat n="advTemp">
      <cat-item lemma="advTemp" tags="ifi.*"/>
    </def-cat>
  </section-def-cats>
  <section-def-attrs>
    <!-- Definição de Atributos -->
    <def-attr n="nbr">
      <attr-item tags="sg"/>
      <attr-item tags="pl"/>
    </def-attr>
    <def-attr n="a_nom">
      <attr-item tags="n"/>
    </def-attr>
    <def-attr n="temps">
      <attr-item tags="pri"/>
      <attr-item tags="fut"/>
      <attr-item tags="past"/>
    </def-attr>
    <def-attr n="pers">
      <attr-item tags="p1"/>
    </def-attr>
    <def-attr n="a_verb">
      <attr-item tags="vblex"/>
    </def-attr>
    <def-attr n="tipus_prn">
      <attr-item tags="prn.subj"/>
      <attr-item tags="prn.obj"/>
    </def-attr>
    <def-attr n="tense">
      <attr-item tags="ifi"/>
      <attr-item tags="past"/>
    </def-attr>
  </section-def-attrs>
  <section-def-vars>
    <!-- Definição de Variáveis -->
    <def-var n="number"/>
  </section-def-vars>
  <section-rules>
    <!-- Definição de Regras -->
    <rule>
      <pattern>
        <pattern-item n="vrb"/>
      </pattern>
      <action>
        <let>
          <clip pos="1" side="t1" part="tense"/>
          <lit-tag v="past"/>
        </let>
      </action>
    </rule>
  </section-rules>
</transfer>
```

```
</let>
<out>
<lu>
  <lit v="advTemp"/>
  <clip pos="1" side="t1" part="temps"/>
</lu>
<b/>
<lu>
  <lit v="prpers"/>
  <lit-tag v="prn"/>
  <lit-tag v="subj"/>
  <clip pos="1" side="t1" part="pers"/>
  <clip pos="1" side="t1" part="nbr"/>
</lu>
<b/>
<lu>
  <clip pos="1" side="t1" part="lem"/>
  <clip pos="1" side="t1" part="a_verb"/>
</lu>
<b/>
</out>
</action>
</rule>
</section-rules>
</transfer>
```