



UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE
CENTRO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA E
DE COMPUTAÇÃO



Uma Solução para Geração Automática de Trilhas em Língua Brasileira de Sinais em Conteúdos Multimídia

Tiago Maritan Ugulino de Araújo

Natal - RN
Setembro de 2012



UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE
CENTRO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA E
DE COMPUTAÇÃO



Uma Solução para Geração Automática de Trilhas em Língua Brasileira de Sinais em Conteúdos Multimídia

Tiago Maritan Ugulino de Araújo

Orientador: Prof. Dr. Guido Lemos S. Filho

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Engenharia Elétrica e de Computação da UFRN (área de concentração: Engenharia de Computação) como parte dos requisitos para obtenção do título de Doutor em Ciências.

Natal, RN, setembro de 2012

A663u Araújo, Tiago Maritan Ugulino de.
Uma solução para geração automática de trilhas em Língua
Brasileira de Sinais em conteúdos multimídia / Tiago Maritan
Ugulino de Araújo.—Natal, RN, 2012.
203f. : il.
Orientador: Guido Lemos S. Filho
Tese (Doutorado) – UFRN/CT
1. Tradução de Linguagem de Programação – Língua
Brasileira de Sinais (LIBRAS). 2. Conteúdos Multimídias.
3. Tecnologias acessíveis – surdos. 4. Síntese de sinais.

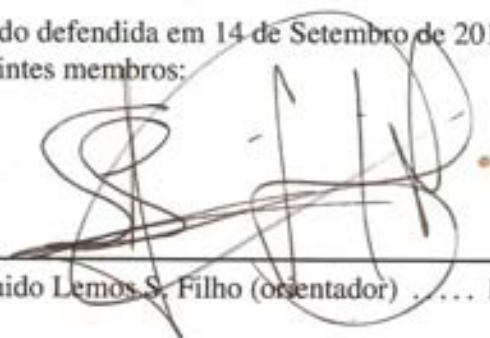
UFPB/BC

CDU: 004.4'4(043)


Uma Solução para Geração Automática de Trilhas em Língua Brasileira de Sinais em Conteúdos Multimídia

Tiago Maritan Ugulino de Araújo

Tese de Doutorado defendida em 14 de Setembro de 2012 pela banca examinadora composta pelos seguintes membros:



Prof. Dr. Guido Lemos S. Filho (orientador) DI/UFPB - PPGEEC/UFRN



Prof. Dr. Luiz Marcos Garcia Gonçalves DCA/UFRN



Prof. Dr. Celso Alberto Saibel Santos DI/UFES



Prof. Dr. Luiz Fernando Gomes Soares DI/PUC-RIO

*Aos meus pais, Mário e Regiane,
meus exemplos de vida, que me
ensinaram desde pequeno a
importância de estudar e batalhar
para conquistar um objetivo na vida.
A eles devo tudo o que sou e o que
aprendi.*

Agradecimentos

À Deus, em primeiro lugar, por ter me dado o dom da vida e ter sempre me dado saúde, paciência, perseverança e fé para enfrentar todos os desafios que tive que enfrentar nesse caminho.

Aos meus pais, Mário e Regiane, pelo amor incondicional, pelo exemplo, incentivo e orientação em todas as decisões de minha vida. Devo a eles tudo o que sou.

À Manu, pelo carinho, companheirismo e paciência, especialmente nos momentos que tive que estar ausente para realizar esse trabalho, e também pelo incentivo incondicional nos momentos de maior desgaste e de desânimo.

À Natália, minha irmã, pelo amor, enorme carinho e incentivo nos momentos desânimo, e especialmente, pela sua paciência em me agüentar nos meus momentos de maior stress.

Ao Prof. Guido, meu orientador, um exemplo de pesquisador e de profissional, uma pessoa que admiro e me espelho profissionalmente. Agradeço demais pela confiança, pelas cobranças realizadas especialmente nas horas mais necessárias, e por tentar sempre tirar o melhor daqueles que o cercam.

À Tatiana, Alexandre, Alisson, Rostand, Raoni, Lacet, Lemos, Danilo, Eduardo, Leonardo Dantas, Leonardo Araújo, Yurika, Hozana, Vandhuy, Luiz, Derzu, Gutenberg, Joãozinho, Andrei, Dênio, Hacks, Igor, Lucenildo, dentre tantos outros colegas de LAViD, pela preciosas dicas, ajudas e valiosas discussões, que foram essenciais no desenvolvimento deste trabalho.

Aos meus amigos e colegas que, embora não tenham sido lembrados diretamente aqui, estão no meu coração. Saibam que até as mais simples palavras de incentivo foram muito importante pra que eu estivesse aqui hoje.

Por fim, agradeço também ao CNPq pelo apoio financeiro para a realização deste trabalho.

Resumo

Os surdos enfrentam sérias dificuldades para acessar informações. As Tecnologias de Informação e Comunicação (TIC) quando são desenvolvidas dificilmente levam em consideração os requisitos específicos destes usuários especiais. O suporte para línguas de sinais, por exemplo, é raramente explorado nessas tecnologias. Além disso, as soluções presentes na literatura relacionadas a tradução automática para línguas de sinais são restritas a um domínio de aplicação específico ou não são viáveis para cenários que necessitam de tradução em tempo real, como, por exemplo, na TV. Para reduzir esses problemas, neste trabalho é proposta uma solução para geração automática de trilhas em Língua Brasileira de Sinais (LIBRAS) em conteúdos digitais multimídia legendados. As trilhas de LIBRAS são geradas a partir de uma estratégia de tradução automática e em tempo real para LIBRAS, que realiza a tradução a partir de fluxos de legendas (como, por exemplo, legendas ou *closed caption*) em língua portuguesa. Além disso, a solução proposta é de domínio geral e dotada de um conjunto de mecanismos que exploram a colaboração e a computação humana para gerar e manter suas construções lingüísticas de forma eficiente. Implementações da solução proposta foram desenvolvidas para as plataformas de TV Digital, Web e Cinema Digital, e um conjunto de experimentos, incluindo testes com usuários surdos, foi desenvolvido para avaliar os principais aspectos da solução. Os resultados mostraram que a solução proposta é eficiente, capaz de gerar e embarcar as trilhas de LIBRAS em cenários que exigem tradução em tempo real, além de ser uma alternativa prática e viável para redução das barreiras de acesso à informação dos surdos, especialmente quando intérpretes humanos não estão disponíveis.

Palavras-chave: Tradução Automática, Língua Brasileira de Sinais, Conteúdos Multimídia, Tecnologias Acessíveis para Surdos, Síntese de Sinais.

Abstract

Deaf people have serious difficulties to access information. The support for sign languages is rarely addressed in Information and Communication Technologies (ICT). Furthermore, in scientific literature, there is a lack of works related to machine translation for sign languages in real-time and open-domain scenarios, such as TV. To minimize these problems, in this work, we propose a solution for automatic generation of Brazilian Sign Language (LIBRAS) video tracks into captioned digital multimedia contents. These tracks are generated from a real-time machine translation strategy, which performs the translation from a Brazilian Portuguese subtitle stream (e.g., a movie subtitle or a closed caption stream). Furthermore, the proposed solution is open-domain and has a set of mechanisms that exploit human computation to generate and maintain their linguistic constructions. Some implementations of the proposed solution were developed for digital TV, Web and Digital Cinema platforms, and a set of experiments with deaf users was developed to evaluate the main aspects of the solution. The results showed that the proposed solution is efficient and able to generate and embed LIBRAS tracks in real-time scenarios and is a practical and feasible alternative to reduce barriers of deaf to access information, especially when human interpreters are not available.

Keywords: Machine Translation, Brazilian Sign Language, Multimedia Contents, Accessibility Technologies for Deaf, Sign Synthesis.

Sumário

Sumário	i
Lista de Figuras	v
Lista de Tabelas	vii
Lista de Símbolos e Abreviaturas	ix
1 Introdução	1
1.1 Motivação	2
1.2 Tese	6
1.3 Hipóteses	7
1.4 Objetivos	8
1.5 Escopo do Trabalho	9
1.6 Organização da Tese	10
2 Fundamentação Teórica	11
2.1 Línguas de Sinais	11
2.1.1 Sinais e seus Parâmetros	12
2.1.2 Espaço de Sinalização	13
2.1.3 Sistema de Representação e Escrita	15
2.1.4 LIBRAS	17
2.2 Sistemas de Tradução Automática	18
2.2.1 Tradução Automática Baseada em Regras	20
2.2.2 Tradução Automática Estatística	22
2.2.3 Tradução Automática Baseada em Exemplos	25
2.3 Computação Humana	27
2.3.1 Jogos com Propósito	29
2.3.2 Projeto reCAPTCHA	30
2.4 Considerações	30

3	Revisão Sistemática da Literatura	33
3.1	Busca, Seleção e Classificação dos artigos	34
3.2	Avaliação Qualitativa das Principais Contribuições	39
3.2.1	Tradução Automática para Língua Indiana de Sinais	40
3.2.2	Tradução Automática para Língua Taiwanesa de Sinais	42
3.2.3	Tradução Automática para Língua Tailandesa de Sinais	44
3.2.4	Tradução Automática para Língua Espanhola de Sinais	45
3.2.5	Tradução Automática para Língua Grega de Sinais	47
3.2.6	Tradução Automática para Língua Japonesa de Sinais	48
3.2.7	Tradução Automática para Língua Americana de Sinais	50
3.2.8	Avaliação dos Principais Resultados	52
3.3	Considerações	52
4	Solução Proposta	55
4.1	Filtragem e Extração de Legendas	56
4.2	Tradução Automática	58
4.3	Animação e Sincronização	60
4.4	Distribuição	63
4.5	Ferramenta de Computação Humana	64
4.5.1	WikiLIBRAS	64
4.5.2	Linguagem de Descrição de Regras de Tradução	66
4.5.3	Linguagem de Descrição de Sinais	68
4.5.4	Modelo de um Agente Animado Virtual-3D	71
4.6	Considerações	72
5	Cenários de Uso	75
5.1	LibrasTV	75
5.1.1	Protocolo de Codificação do LibrasTV	78
5.1.2	Implementação dos Componentes	81
5.2	LibrasWeb	86
5.2.1	Implementação dos Componentes	86
5.2.2	Integração na Nuvem	88
5.3	CineLibras	91
5.4	WikiLIBRAS	93
5.5	Testes Preliminares com os Protótipos	96
5.5.1	LibrasTV	96

5.5.2	LibrasWeb	101
5.6	Considerações	104
6	Experimentos	105
6.1	Execução do Experimento e Análise dos Dados	106
6.1.1	Avaliação do Nível de Compreensão dos Conteúdos	106
6.1.2	Avaliação do Desempenho de Tradução	113
6.1.3	Avaliação da Ferramenta de Computação Humana	115
6.2	Considerações	120
7	Conclusões	123
7.1	Principais Contribuições	124
7.1.1	Contribuições Científicas Gerais	124
7.1.2	Contribuições Colaterais	126
7.1.3	Produção Técnica-Científica	127
7.2	Propostas de Trabalhos Futuros	130
	Referências bibliográficas	132
A	Protocolo da Revisão Sistemática	143
B	Artigos da Revisão Sistemática	149
C	Parâmetros da LDS	155
D	Planejamento dos Experimentos	157
D.1	Definição do Experimento	157
D.1.1	Avaliação do Nível de Compreensão dos Conteúdos	157
D.1.2	Avaliação do Desempenho de Tradução	158
D.1.3	Avaliação da Ferramenta de Computação Humana	159
D.2	Planejamento do Experimento	160
D.2.1	Avaliação do Nível de Compreensão dos Conteúdos	160
D.2.2	Avaliação do Desempenho de Tradução	164
D.2.3	Avaliação da Ferramenta de Computação Humana	166
E	Questionários	171
E.1	Questionário 1: Avaliação dos Conteúdos Acessíveis	171
E.2	Questionário 2: Avaliação do WikiLIBRAS	175

Lista de Figuras

2.1	Configurações de mão da LIBRAS	14
2.2	Diagrama do espaço de sinalização	15
2.3	Representação do sinal "don't know" na notação de Stokoe	16
2.4	Representação do sinal "nineteen" na notação HamNoSys	16
2.5	Representação do sinal "deaf" na notação SignWriting	16
2.6	Três tipos de desafios linguísticos dos sistemas de tradução automática.	19
2.7	Tipos de sistemas de tradução automática baseado em regras	20
2.8	Exemplo de reCAPTCHA.	31
4.1	Visão esquemática da solução proposta	57
4.2	Visão esquemática do componente de Tradução Automática	59
4.3	Visão esquemática do componente de Animação	62
4.4	Visão esquemática da ferramenta WikiLIBRAS	65
4.5	Exemplo de representação da regra de topicalização: uma sentença na ordem sujeito+verbo+objeto é traduzida como objeto+sujeito+verbo.	68
4.6	Exemplo de representação do sinal LÁBIO	70
4.7	Exemplo de representação do sinal CORRETO.	70
4.8	(a) Modelo do avatar-3D. Ênfase nos ossos da (b) face, (c) das mãos e (d) do corpo.	72
5.1	Sistema de TV Digital.	76
5.2	Arquitetura geral do protótipo LibrasTV.	78
5.3	Diagrama de classe do componente de Tradução Automática	82
5.4	Diagrama de classes da aplicação LibrasTV.	84
5.5	Capturas de tela da execução da aplicação LibrasTV sobre o Openinga	85
5.6	Arquitetura geral de integração do LibrasWeb na nuvem	86
5.7	Capturas de tela do vídeo gerado pelo LibrasWeb	88
5.8	Visão esquemática da integração do LibrasWeb em uma infraestrutura de computação em nuvem.	89

5.9	Capturas de tela do protótipo do LibrasWeb: (a) tela de seleção da posição da janela de LIBRAS; (b) tela de apresentação da cópia acessível do conteúdo.	90
5.10	Arquitetura geral de integração do CineLibras	91
5.11	Demonstração do CineLIBRAS no XIII WRNP. O vídeo do Cinema aparece ao fundo, enquanto que o vídeo de LIBRAS é gerado pelo protótipo e transmitido para o dispositivo móvel do usuário.	92
5.12	Captura da tela principal de cadastro de um sinal no WikiLIBRAS.	93
5.13	Capturas de tela para configuração dos fonemas: (a) expressão facial, (b) configuração de mão, (c) orientação da palma da mão e (d) ponto de articulação.	94
5.14	Capturas de tela para (a) Cadastro e (b) Edição de Regras.	95
5.15	Captura de tela do vídeo do sinal gerado pelo WikiLIBRAS.	96
5.16	Largura de banda (em Kbps) utilizada pelo fluxo de LIBRAS codificado .	101
5.17	Largura de banda (em Kbps) utilizada pelo fluxo de vídeo com intérprete de LIBRAS	102
5.18	Tempo médio de resposta para diferentes requisições nas duas abordagens.	103
5.19	Custo médio para diferentes requisições nas duas abordagens.	103
6.1	Princípios básicos de um experimento	105
6.2	Grau de Escolaridade dos usuários do Teste	107
6.3	Algumas fotos do experimento com usuários (nível de compreensão) (a) Alguns usuários do experimento, (b) Fase de preparação dos usuários, (c) Usuários avaliando conteúdos legendados, (d) Usuários avaliando conteúdos com trilhas de LIBRAS geradas pela solução proposta.	109
6.4	Resultados dos testes de compreensão dos conteúdos	110
6.5	Gráfico de caixa (<i>box plot</i>) para os testes de compreensão dos conteúdos .	111
6.6	Grau de Escolaridade dos usuários do Teste	116
6.7	Algumas fotos do experimento com usuários (WikiLIBRAS).	116
6.8	Taxa de sinais gerados corretamente pelos usuários (WikiLIBRAS vs Manual).	117
6.9	Tempo médio para gerar os sinais corretamente (WikiLIBRAS vs Manual).	118

Lista de Tabelas

3.1	Resultados do processo de busca e seleção de artigos	35
3.2	Classificação dos artigos de acordo com as categorias	36
3.3	Classificação dos artigos de acordo com os métodos de Tradução Auto- mática	37
3.4	Tipo de avaliação utilizada nos trabalhos de domínio geral (33 artigos) . .	39
3.5	Artigos que envolvem avaliação com usuários	41
3.6	Artigos que envolvem avaliação com usuários	53
4.1	Classes morfo-sintática da Linguagem de Descrição de Regras de Tradução	67
5.1	Sintaxe das mensagens LCM	79
5.2	Sintaxe das mensagens LDM	79
5.3	Valores do campo resolution.	80
5.4	Sintaxe dos eventos DSM-CC	81
5.5	Fluxos de Vídeo MPEG-2 TS com legendas utilizados nos testes	97
5.6	Medidas do percentual de classificação correta das classificações morfo- sintáticas sobre o corpus "Bosque"	98
5.7	Valores de BLEU e WER para o LibrasTV e para uma solução baseada em Tradução Direta.	99
5.8	Medidas do tempo (atraso) médio de cada componente do LibrasTV . . .	100
6.1	Nível de conhecimento em Português e LIBRAS dos Usuários	108
6.2	Resultados do teste-t para a Hipótese 1 (compreensão dos conteúdos) . . .	110
6.3	Valores médios para os aspectos avaliados (escalas de 1 a 6)	112
6.4	Resultados do teste-t para a Hipótese 2 (adequação a gramática de LIBRAS)	113
6.5	Medidas do desempenho de tradução para os componentes do protótipo LibrasTV	114
6.6	Resultados do teste-t para a Hipótese desse experimento (desempenho ou atraso da tradução)	115
6.7	Nível de conhecimento em Português e LIBRAS dos Usuários	115

6.8	Resultados do teste-t para a Hipótese desse experimento (tempo médio para gerar os sinais)	119
6.9	Valores médios para os aspectos avaliados (escalas de 1 a 6)	120
A.1	Termos de busca	144
C.1	Tipos de trajetórias de movimento	155
C.2	Ponto de Articulação	155
C.3	Ponto de Articulação	156
C.4	Expressões faciais	156
D.1	Conteúdos multimídia utilizados nos testes	163
D.2	Sinais utilizados nos testes	169

Lista de Símbolos e Abreviaturas

AaaS: *Accessibility as a Service*

Amazon EC2: *Amazon Elastic Compute Cloud*

ASL: *American Sign Language*

ATM: *Automatic Teller Machine*

BE: *Base de Exemplos*

BLEU: *Bilingual Evaluation Understudy*

BSL: *British Sign Language*

CSL: *Chinese Sign Language*

DCP: *Digital Cinema Package*

DSM-CC: *Digital Storage Media-Command and Control*

EBMT: *Example-Based Machine Translation*

FSL: *French Sign Language*

GSL: *Greek Sign Language*

GWAP: *Game With a Purpose*

HamNoSys: *Hamburg Notation System*

HTTP: *HyperText Transfer Protocol*

IA: *Inteligência Artificial*

IaaS: *Infrastructure as a Service*

IBGE: *Instituto Brasileiro de Geografia e Estatística*

IGDA: *International Game Design Association*

IK: *Inverse Kinematics*

IR: *Intermediate Representation*

IrishSL: *Irish Sign Language*

ISL: *Irish Sign Language*

JMF: *Java Media Framework*

JSL: *Japanese Sign Language*

KSL: *Korean Sign Language*

LCM: LIBRASControlMessage

LDM: LIBRASDataMessage

LDRT: Linguagem de Descrição de Regras de Tradução

LDS: Linguagem de Descrição de Sinais

LFG: *Lexical Functional Grammars*

LIBRAS: Língua Brasileira de Sinais

LS: Línguas de Sinais

LSE: Língua Espanhola de Sinais

MT: *Machine Translation*

OCR: *Optical Character Recognition*

OMS: Organização Mundial de Saúde

PCFG: *Probabilistic Context-Free Grammars*

PCR: *Program Clock Reference*

PER: *Position Independent Error Rate*

PLN: Processamento de Linguagem Natural

PPM: *Prediction by Partial Matching*

PTS: *Presentation Timestamp*

QP: Questão de Pesquisa

RBMT: *Rule-Based Machine Translation*

RNP: Rede Nacional de Ensino e Pesquisa

SBRC: Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos

SBTVD: Sistema Brasileiro de TV Digital

SCFG: *Synchronous Context-Free Grammars*

SER: *Sign Error Rate*

SMT: *Statistical Machine Translation*

ThaiSL: *Thai Sign Language*

TIC: Tecnologia da Informação e Comunicação

TS: *Transport Stream*

TSL: *Taiwanese Sign Language*

UDP: *User Datagram Protocol*

UFPB: Universidade Federal da Paraíba

USB: *Universal Serial Bus*

WER: *Word Error Rate*

WRNP: Workshop da RNP

Capítulo 1

Introdução

A comunicação é essencial na interação humana. Segundo Russell & Norvig (2004), "a comunicação é a troca intencional de informações provocada pela produção e percepção de sinais extraídos de um sistema compartilhado de sinais convencionais". Através desses sistemas compartilhados de sinais, denominados línguas, os seres humanos podem comunicar suas idéias, sentimentos e pensamentos, registrar seus conhecimentos, transmitir seus padrões culturais, dentre outros.

A língua utilizada por um indivíduo para se comunicar depende do grupo em que ele está inserido. Os ouvintes, por exemplo, se comunicam naturalmente através de línguas orais, isto é, através de sons articulados que são percebidos pelo sistema auditivo. Os surdos, por outro lado, se comunicam naturalmente através de línguas gestuais, também denominadas línguas de sinais. Nessa modalidade, elementos lingüísticos manuais, corporais e faciais são utilizados para articular os sinais que são compreendidos através do sistema visual. Isso implica que para os surdos, as línguas orais, utilizadas cotidianamente pela maioria das pessoas e em praticamente todos os meios de comunicação, representam "uma segunda língua"[Campos & Girafa 2000].

Segundo o censo demográfico do Instituto Brasileiro de Geografia e Estatística (IBGE) de 2010, existem, no Brasil, cerca de 9,7 milhões de pessoas com algum nível de deficiência auditiva, o que representa cerca de 5,1% da população brasileira [IBGE 2010]. Em termos mundiais, a estimativa da Organização Mundial de Saúde (OMS) é de que aproximadamente 275 milhões de pessoas apresentem algum nível de deficiência auditiva [WHO 2012]. Isso implica que os surdos representam uma parcela significativa da população brasileira e mundial.

O presente trabalho aborda uma solução para geração automática de trilhas de Língua Brasileira de Sinais (LIBRAS) em conteúdos multimídia com o objetivo de melhorar o acesso dos surdos às Tecnologias da Informação e Comunicação (TIC) e, conseqüentemente, reduzir as barreiras de comunicação e de acesso à informação dessas pessoas.

1.1 Motivação

As línguas de sinais (LS) são a forma natural de comunicação entre os surdos. Diferentemente das línguas orais que utilizam o som na comunicação, as línguas de sinais utilizam um canal visual, isto é, um conjunto de elementos lingüísticos manuais, corporais e faciais para articular os sinais [Goes 1996]. O emissor constrói uma sentença a partir desses elementos lingüísticos e o receptor utiliza o sistema visual para compreender o que está sendo comunicado. Desta forma, os sinais são construídos de acordo com as possibilidades perceptivas do sistema visual humano [Macedo 1999].

Como as línguas orais representam para os surdos apenas uma segunda língua, muitos deles têm dificuldade de compreender e se comunicar através de textos em línguas orais. Uma vez que essas línguas possuem uma grafia baseada em sons, a maioria dos surdos passa vários anos na escola e não conseguem aprender a ler e escrever na língua oral de seu país [Stumpf 2000]. No Brasil, por exemplo, segundo o censo demográfico do IBGE de 2000 [IBGE 2000], cerca de 97% dos surdos não concluem o ensino médio¹. Um outro exemplo que pode ser citado é que um estudo realizado por Wauters (2005) com crianças e adolescentes surdos holandeses de 7 a 20 anos de idade, mostrou que apenas 25% deles possuem uma capacidade de leitura igual ou superior ao de uma criança sem deficiência de 9 anos.

Além disso, as Tecnologias da Informação e Comunicação (TIC) quando são desenvolvidas raramente levam em conta os requisitos e necessidades dos portadores de necessidades especiais [Haddon & Paul 2001]. O suporte para línguas de sinais, por exemplo, é raramente explorado nessas tecnologias. Na TV, por exemplo, o suporte a línguas de sinais é, em geral, limitado a uma janela com um intérprete de língua de sinais, apresentada juntamente com o vídeo original do programa (*wipe*). Essa solução além de possuir altos custos operacionais para geração e produção (câmeras, estúdio, equipe, etc.) dos conteúdos, necessita de intérpretes humanos em tempo integral, o que acaba restringindo seu uso a uma pequena parcela da programação. Essas dificuldades resultam em uma grande barreira para a comunicação com outras pessoas, o acesso a informações, a aquisição de conhecimentos, dentre outros. Essas barreiras motivam a primeira questão de pesquisa

¹No censo demográfico do IBGE de 2010 [IBGE 2010], foi identificado que cerca 61% dos portadores de deficiência (sem distinção do tipo de deficiência) não possuem qualquer instrução ou possuem ensino médio incompleto e cerca de 75% não conclui o ensino médio. Contudo, o censo não disponibiliza essas informações específicas para os surdos.

(QP) deste trabalho:

- **QP1:** *Como é possível reduzir as barreiras de acesso à informação das pessoas surdas nas TIC, especialmente quando intérpretes humanos não estão disponíveis?*

Na literatura científica, existem alguns trabalhos direcionados para as necessidades comunicativas dos surdos [Lee et al. 2007][Lee et al. 2005][Starner et al. 1998]. Esses trabalhos oferecem soluções tecnológicas para atividades cotidianas que permitem que pessoas com necessidades especiais assistam e compreendam televisão, interajam com outras pessoas ou escrevam uma carta. O uso de legendas com informações emotivas em filmes e programas de televisão [Lee et al. 2007] e o desenvolvimento de jogos para crianças surdas [Lee et al. 2005] são exemplos desse tipo de solução.

Outros trabalhos são relacionados ao uso de estratégias de tradução automática para línguas de sinais [Gallo et al. 2009][Morrissey 2008][Othman & Jemni 2011][San-Segundo et al. 2006][San-segundo et al. 2008a][San-segundo et al. 2008b][San-segundo et al. 2011][Veale et al. 1998][Zhao et al. 2000]. Considerando que as línguas de sinais possuem gramáticas próprias, com regras específicas em seus níveis lingüísticos, morfológico e sintático, um processo de tradução direta da língua oral para língua de sinais, ou seja, sem utilização de processamento, interpretação ou análise de contexto, não é apropriado. Além disso, os dois tipos de línguas (orais e de sinais) se diferenciam quanto a sua estrutura de desenvolvimento ao longo do tempo. Enquanto as línguas orais possuem uma estrutura seqüencial, ou seja, os fonemas são produzidos seqüencialmente no tempo, as línguas de sinais possuem uma estrutura paralela, podendo emitir sinais que envolvem simultaneamente diversas partes do corpo do sinalizador [Brito 1995][Quadros 1997]. Em virtude disso, um sistema de tradução automática entre os dois tipos de línguas deve levar em consideração, tanto as mudanças referentes ao canal de comunicação utilizado (isto é, mudança do canal sonoro das línguas orais para o canal visual das línguas de sinais), quanto às mudanças gramaticais e estruturais (isto é, mudança de estrutura seqüencial das línguas orais para a estrutura paralela das línguas de sinais).

Veale et al. (1998), por exemplo, propôs um sistema de tradução automática multilíngue para traduzir textos em inglês para língua americana de sinais (*American Sign Language - ASL*), língua irlandesa de sinais (*Irish Sign Language - IrishSL*) e língua japonesa de sinais (*Japanese Sign Language - JSL*). Esse sistema é baseado numa arquitetura de quadro negro (*blackboard control architecture*) [Othman & Jemni 2011] e possui um conjunto de agentes que cooperam para gerar os conteúdos traduzidos. Esse trabalho explora e estende alguns conceitos de Inteligência Artificial (IA) para línguas de sinais como, por exemplo, representação do conhecimento, raciocínio metafórico, arquiteturas baseadas

em quadro negro [Morrissey 2008], mas nenhum teste ou experimento foi realizado para avaliar a viabilidade e qualidade solução.

Zhao et al. (2000) propuseram uma abordagem baseada em interlíngua para tradução de textos em inglês para ASL. Nessa solução, os dados de entrada são analisados e uma representação intermediária (*intermediate representation* - IR) é gerada a partir da análise desses dados. Um sintetizador então utiliza essa representação intermediária para gerar os sinais. No entanto, da mesma forma, que no trabalho proposto por Veale et al. (1998), nenhum teste ou experimento foi realizado para avaliar a solução. Othman & Jemni (2011) propuseram uma estratégia para alinhamento de palavras e incluíram-na em tradutor estatístico de inglês para ASL. No entanto, apenas a estratégia de alinhamento de palavras foi avaliada, não sendo realizado nenhum teste para avaliar a qualidade e velocidade da tradução, por exemplo.

Gallo et al. (2009), San-Segundo et al. (2006), San-segundo et al. (2008a), San-segundo et al. (2008b), San-segundo et al. (2011) propuseram uma arquitetura para traduzir voz em língua espanhola de sinais (LSE) com enfoque em ajudar pessoas surdas no atendimento em serviços públicos como, por exemplo, na renovação da carteira nacional de habilitação. A idéia do sistema é facilitar o diálogo entre surdos e prestadores de serviços públicos nesse tipo de serviço. No entanto, essa solução também é restrita a um domínio de aplicação específico (atendimento em serviços públicos). Além disso, o tempo médio reportado para traduzir cada sentença de voz para LSE foi de cerca de 8 segundos, o que torna a solução inviável em cenários que exigem tradução em tempo real, como, por exemplo, na TV.

Existem na literatura científica outros trabalhos relacionados a tradução automática para línguas de sinais (esses trabalhos serão apresentados em detalhes no Capítulo 3), mas, em geral, esses trabalhos também não avaliaram a viabilidade e qualidade da solução [Othman & Jemni 2011][Veale et al. 1998][Zhao et al. 2000], ou são restritos a um domínio de aplicação específico, ou não são viáveis para cenários que necessitam de tradução em tempo real [Gallo et al. 2009] [San-Segundo et al. 2006], [San-segundo et al. 2008a], [San-segundo et al. 2008b], [San-segundo et al. 2011], como, por exemplo, na TV. Essas limitações motivam a segunda questão de pesquisa deste trabalho:

- **QP2:** *É possível desenvolver um tradutor automático para línguas de sinais que seja eficaz em cenários que exigem tradução em tempo real a partir de fluxos de legendas, como, por exemplo, na TV, e que também seja de domínio geral?*

Outra dificuldade encontrada no desenvolvimento desses tradutores é que o desenvolvimento das suas construções lingüísticas (regras de tradução, *corpus* bilingüe, dicionário

de sinais, etc.) é, em geral, uma tarefa não-trivial e que demanda muito trabalho manual. Além disso, considerando que as línguas de sinais são línguas naturais e vivas, novos sinais ou novas construções gramaticais podem surgir espontaneamente ao longo do tempo. Isso implica que esses novos sinais e construções também precisam ser incluídos na solução, caso contrário, a qualidade dos conteúdos gerados por ela tenderia a piorar ao longo do tempo, tornando-a desatualizada.

Uma alternativa para acelerar o processo de desenvolvimento e manutenção² dessas construções lingüísticas seria desenvolver e incorporar uma infraestrutura ou um conjunto de ferramentas que explorassem a colaboração ou computação humana³ para geração semi-automatizada dessas construções. Dessa forma, seria possível que colaboradores humanos auxiliassem no desenvolvimento e manutenção dessas construções lingüísticas, acelerando o processo de desenvolvimento desses tradutores. Isso implicaria que também seria possível atualizar o conjunto de sinais e/ou regras gramaticais sempre que necessário, permitindo, dessa forma, que os conteúdos acessíveis gerados pela solução evoluam ao longo do tempo. Contudo, a forma como essas ferramentas de colaboração/computação humana podem ser desenvolvidas e incorporadas nos tradutores automáticos para línguas de sinais motivam a terceira questão de pesquisa deste trabalho:

- **QP3:** *Como é possível acoplar ferramentas de colaboração e computação humana nos tradutores automáticos para línguas de sinais de forma que colaboradores humanos auxiliem no processo de geração desses conteúdos e melhorem a qualidade dos conteúdos acessíveis gerados ao longo do tempo?*

Cientificamente, a principal motivação deste trabalho é abordar essas três questões de pesquisa levantadas acima e propor soluções computacionais dentro da perspectiva da pesquisa abordada. Mais especificamente, a motivação é reduzir as barreiras de comunicação e acesso à informação dos surdos nas TIC, desenvolvendo uma solução para geração automática de trilhas em Língua Brasileira de Sinais (LIBRAS) em conteúdos digitais multimídia. LIBRAS é a língua de sinais utilizada pela maioria dos surdos brasileiros e reconhecida pela Lei Brasileira nº 10.436, de 24 de abril de 2002, e a geração de trilhas em LIBRAS deve ser baseada em uma estratégia de tradução automática e em tempo real para LIBRAS. Além disso, a solução deve ser de domínio geral e deve ser dotada de mecanismos que permitam explorar a colaboração ou computação humana (*human computation*) para permitir a geração e manutenção das construções lingüísticas, acelerando

²isto é, permitir que novas construções sejam desenvolvidas ou as atuais sejam editadas

³De acordo com Quinn & Bederson (2011), a computação humana pode ser definida como um paradigma que utiliza o poder de processamento humano para resolver problemas que os computadores ainda não são capazes de resolver sozinhos.

o processo de desenvolvimento da solução, e permitindo a melhoria da qualidade dos conteúdos acessíveis gerados ao longo do tempo.

Do ponto de vista tecnológico, a motivação inicial é modelar um agente animado virtual (avatar), desenvolver dicionários em LIBRAS, e implementar protótipos dessa solução em diferentes TIC, como, por exemplo, TV digital, Cinema Digital e Web. Por fim, do ponto de vista social, a motivação inicial é reduzir as barreiras de comunicação e acesso à informação dos surdos, através do acesso a essas tecnologias.

1.2 Tese

A tese defendida neste trabalho é de que é possível gerar automaticamente trilhas (ou vídeos) de LIBRAS em conteúdos digitais multimídia legendados, a partir de uma estratégia de tradução automática e em tempo real de fluxos de legendas (como, por exemplo, legendas de filmes ou fluxos de *closed caption*) em língua portuguesa. Além disso, a solução proposta deve ser de domínio geral e dotada de mecanismos que permitam explorar a colaboração e a computação humana para gerar e manter as construções lingüísticas e melhorar a qualidade dos conteúdos acessíveis gerados ao longo do tempo.

A solução também deve utilizar um conjunto de estratégias eficientes, como, por exemplo, uma etapa de tradução textual de língua portuguesa para *glosa* (uma representação textual de LIBRAS) que consome pouco tempo computacional e dicionários de LIBRAS para evitar a renderização dos sinais em tempo real, reduzindo os recursos computacionais necessários para gerar as trilhas em LIBRAS em tempo real. Adicionalmente, os dicionários de LIBRAS podem estar armazenados nos clientes ou serem carregados pelo canal de comunicação, permitindo que as diferenças regionais presentes nas línguas de sinais sejam respeitadas.

Além disso, a estratégia de colaboração e computação humana inclui uma ferramenta de colaboração, denominada WikiLIBRAS, linguagens de descrição de sinais e de regras de tradução e um modelo 3D que permite que usuários colaboradores possam gerar semi-automaticamente e manter as construções lingüísticas (dicionário de sinais e regras de tradução) da solução.

Como cenários de uso para esta tese, implementações da solução proposta foram desenvolvidas para o sistema brasileiro de televisão digital (SBTVD), para plataformas Web e para Cinema Digital, respectivamente. As arquiteturas e adaptações necessárias desses cenários também são contribuições deste trabalho. Além disso, um processo de experimentação quantitativo e qualitativo foi conduzido com usuários surdos para avaliar a solução e testar as hipóteses.

1.3 Hipóteses

Com base nas questões de pesquisa apresentadas na Seção 1.1 e na Tese proposta na Seção 1.2, três hipóteses gerais foram definidas para o desenvolvimento do trabalho. Informalmente, essas hipóteses podem ser definidas da seguinte forma:

1. É possível reduzir as barreiras de acesso à informação dos surdos brasileiros nas TIC através da geração automática de trilhas de LIBRAS em conteúdos multimídia a partir da tradução automática de fluxos de legendas.
2. Os conteúdos acessíveis gerados pela solução serão produzidos de forma eficiente, inclusive em cenários que exijam tradução em tempo real e que os conteúdos sejam de domínio geral (como, por exemplo, na TV).
3. A computação humana poderá ser integrada e explorada na solução. Ela permitirá que especialistas em LIBRAS (incluindo surdos) gerem as construções lingüísticas da solução semi-automaticamente, aumentando a produtividade com relação a geração manual dessas construções.

Formalmente, essas hipóteses podem ser definidas da seguinte forma:

1. **Hipótese nula H_0 :** Os conteúdos acessíveis gerados pela solução proposta não melhorarão o nível de compreensão dos usuários surdos brasileiros aos conteúdos multimídia quando confrontados, por exemplo, com conteúdos com legendas.
Hipótese alternativa H_1 : Os conteúdos acessíveis gerados pela solução serão melhor compreendidos pelos surdos com relação, por exemplo, a conteúdos com legendas, reduzindo suas barreiras de acesso à informação nas TIC.
2. **Hipótese nula H_0 :** A solução não será eficiente e, conseqüentemente, não será capaz de gerar os conteúdos acessíveis de forma eficiente em cenários que exigem tradução em tempo real e de domínio geral, onde a medida de eficiência será o atraso médio de tradução produzido pelos componentes da solução.
Hipótese alternativa H_1 : A solução será eficiente e capaz de traduzir conteúdos em tempo real e de domínio geral.
3. **Hipótese nula H_0 :** As estratégias de colaboração/computação humana não poderão ser integradas na solução proposta ou não aumentarão a produtividade na geração das construções lingüísticas da solução, quando confrontadas com a produtividade de especialistas desenvolvendo essas construções manualmente. A medida de produtividade será o tempo médio gasto para desenvolver essas construções lingüísticas corretamente.

Hipótese alternativa H_1 : Essas estratégias poderão ser integradas na solução e aumentarão a produtividade na geração das construções lingüísticas da solução com relação a geração manual dessas construções.

No Capítulo 6 será apresentado um projeto de experimentos desenvolvido para testar essas hipóteses. Esse projeto inclui a definição, o planejamento, a execução do experimento, além da formulação de um conjunto de hipóteses mais específicas (derivadas dessas hipóteses gerais) para testar e avaliar as hipóteses gerais.

1.4 Objetivos

O objetivo geral desse trabalho é propor um sistema que permita que trilhas de LIBRAS sejam geradas automaticamente e em tempo real em conteúdos multimídia a partir de trilhas de legendas em língua portuguesa, especialmente quando intérpretes de LIBRAS não estão disponíveis. As trilhas de LIBRAS serão geradas utilizando representações dos sinais da língua usando um modelo-3D. Além disso, a solução deve ser dotada de um conjunto de estratégias para permitir que colaboradores humanos melhorem a qualidade dos conteúdos produzidos, gerando novos sinais e regras de tradução ou editando os sinais e regras já existentes.

Para que o objetivo geral desse trabalho seja alcançado, faz-se necessário que sejam atendidos os seguintes objetivos específicos:

1. Levantamento bibliográfico detalhado sobre as principais estratégias relacionadas a tradução automática para línguas de sinais;
2. Projeto e desenvolvimento de uma estratégia de tradução automática e em tempo real de textos em língua portuguesa para glosa;
3. Modelagem de um agente animado virtual 3D (avatar-3D) com capacidade de realizar movimentos corporais, manuais e faciais;
4. Modelagem e desenvolvimento de um dicionário de sinais multimídia que armazenará uma representação de todos os sinais em LIBRAS com base no modelo do avatar-3D;
5. Desenvolvimento de um componente de síntese de sinais, responsável por converter a glosa em um vídeo de apresentação de LIBRAS com base no dicionário de sinais;
6. Especificação de uma linguagem para descrição de sinais em LIBRAS que permita descrever sinais em LIBRAS com base no modelo do avatar-3D;

7. Especificação de uma linguagem para descrição de regras de tradução que permita descrever regras de tradução a serem aplicadas pela estratégia de tradução de textos em língua portuguesa para glosa;
8. Projeto e desenvolvimento de uma ferramenta de colaboração que permita que sinais e regras de tradução sejam definidas de acordo com as linguagens de descrição de sinais e de regras de tradução, respectivamente;
9. Implementação da solução proposta e integração nas plataformas de TV Digital, Web e Cinema Digital;
10. Aplicação de um processo de experimentação qualitativo e quantitativo com usuários surdos brasileiros para testar as hipóteses.

1.5 Escopo do Trabalho

O problema abordado neste trabalho possui natureza fortemente interdisciplinar, envolvendo as áreas de Acessibilidade, Computação Gráfica, Inteligência Artificial, Linguística, dentre outras. Dessa forma, o tratamento pleno de todos os aspectos do problema não é trivial. Portanto, restringimos o problema a um escopo tratável onde foram considerados os seguintes aspectos:

1. LIBRAS é a língua de sinais alvo do trabalho. Dessa forma, a proposta neste trabalho não envolve gerar automaticamente conteúdos para todas as línguas de sinais. O foco é gerar conteúdos em LIBRAS. No entanto, a idéia é que a proposta seja facilmente adaptada para outras línguas de sinais;
2. Apenas mídias textuais serão usadas como base para geração dos conteúdos acessíveis. Os conteúdos acessíveis serão gerados apenas a partir de mídias baseadas em texto como legendas, *closed caption*, entre outros. Isso implica que a solução proposta neste trabalho não contempla a geração de conteúdos acessíveis a partir de mídias de áudio (fluxos de voz ou de áudio).
3. As informações não literais presentes nas legendas ocultas (*closed caption*), como, por exemplo, músicas, trovoadas, rangidos, gargalhadas, dentre outros, embora sejam semanticamente importantes na transmissão dos conteúdos, não são tratadas neste trabalho. Incluir o suporte para este tipo de informação na estratégia de tradução é uma proposta de trabalho futuro.
4. A proposta do trabalho não é substituir ou se equiparar aos intérpretes humanos. Atualmente, a qualidade das técnicas de tradução automática e de sinalização vir-

tual⁴ ainda não se aproximam da qualidade de tradução e sinalização humana. Dentre as razões para esse distanciamento pode-se destacar a dificuldade dos sistemas de tradução automática em compreender profundamente os conteúdos que estão sendo comunicados, assim como a dificuldade das estratégias de sinalização virtual de representar emoções e movimentos de forma tão natural quanto os humanos. Em consequência disso, a proposta não é substituir intérpretes humanos ou gerar traduções tão boas quanto as geradas por intérpretes humanos. A proposta é desenvolver uma solução complementar, prática, eficiente e de baixo custo que possa ser utilizada para reduzir as barreiras de comunicação e acesso a informação dos surdos nas TIC, especialmente quando intérpretes humanos não estão disponíveis ou não são viáveis.

1.6 Organização da Tese

Esse trabalho está estruturado em sete capítulos. Este primeiro capítulo apresenta a motivação do trabalho, introduzindo conceitos relevantes como línguas de sinais e sua utilização e importância na comunicação dos surdos. Além disso, a tese também é apresentada, juntamente as hipóteses gerais, o escopo e os objetivos do trabalho.

O segundo capítulo apresenta uma fundamentação teórica detalhada sobre os conceitos relevantes para compreensão do trabalho, como, por exemplo, línguas de sinais, LIBRAS, sistemas de tradução automática entre línguas orais e computação humana. No terceiro capítulo é apresentada uma revisão sistemática da literatura sobre sistemas de tradução automática para línguas de sinais. O quarto capítulo apresenta uma descrição detalhada da solução proposta neste trabalho, seus principais componentes e as estratégias adotadas para desenvolver esses componentes.

O quinto capítulo apresenta implementações, integrações e adaptações da solução proposta neste trabalho em diferentes TIC como, por exemplo, TV Digital, Web e Cinema Digital. No sexto capítulo é apresentado o processo de experimentação utilizado para avaliar a solução proposta. Esse processo inclui o projeto (ou planejamento) dos experimentos que define: as hipóteses, sujeitos, objetos, variáveis, testes estatísticos e instrumentos; a operação do experimento (isto é, como o experimento foi executado), além de analisar os dados e discutir os resultados.

Por fim, no sétimo capítulo são apresentadas as considerações finais, contribuições e algumas propostas de trabalhos futuros.

⁴Neste trabalho, o termo sinalização virtual é usado para representar as estratégias que utilizam agentes animados virtuais (isto é, avatares) para representação dos sinais em línguas de sinais

Capítulo 2

Fundamentação Teórica

Nesse capítulo serão apresentados os principais conceitos que fundamentam este trabalho. Inicialmente, serão expostas as principais características, propriedades e conceitos relacionados às línguas de sinais, especialmente a língua brasileira de sinais (LIBRAS). Em seguida, os sistemas de tradução automática serão apresentados, destacando as principais estratégias e métodos utilizados. Por fim, os principais conceitos relacionados a computação humana serão apresentados.

2.1 Línguas de Sinais

Conforme mencionado anteriormente, os surdos se comunicam naturalmente através das línguas de sinais. Segundo Brito (1995), elas são consideradas línguas naturais, pois surgem espontaneamente da interação entre os deficientes auditivos e podem expressar qualquer conceito descritivo, concreto, racional, literal, metafórico, emocional ou abstrato.

Normalmente, algumas pessoas possuem uma visão equivocada sobre as línguas de sinais, considerando-as como uma interpretação das línguas orais. No entanto, elas são línguas próprias e contêm toda uma estrutura lingüística específica, assim como as línguas orais [Brito 1995][Quadros 1997].

Elas possuem uma natureza gestual-visual, visto que o emissor da informação constrói a sentença a partir de elementos lingüísticos manuais, corporais e faciais e o receptor percebe e compreende essas informações através do sistema visual. Elas também possuem gramáticas próprias e são compostas dos diversos níveis lingüísticos, como morfologia, sintaxe e semântica [Brito 1995]. De forma similar às línguas orais, elas também possuem itens léxicos que são denominados sinais [Stokoe 1980].

Existem, no mundo, diversas línguas de sinais, cada uma contendo suas próprias regras gramaticais, vocabulários e fonemas [Buttussi et al. 2007]. A língua americana

de sinais (*American Sign Language* - ASL) [Lee et al. 2005][Stokoe 1980][Veale et al. 1998][Zhao et al. 2000], a língua britânica de sinais (*British Sign Language* - BSL) [Stokoe 1980], a língua espanhola de sinais (LSE) [Gallo et al. 2009][Lopez-Ludena et al. 2011][San-Segundo et al. 2006] e a língua francesa de sinais de sinais (*French Sign Language* - FSL) [Stokoe 1980] são exemplos de línguas de sinais utilizadas, respectivamente, nos Estados Unidos, Inglaterra, Espanha e França. No Brasil, a língua de sinais utilizada pela maioria dos surdos brasileiros e reconhecida pela lei brasileira nº 10.436, de 24 de abril de 2002, é a língua brasileira de sinais representada pela sigla LIBRAS [Brito 1995][Campos & Girafa 2000][Goes 1996][Felipe 2007]. Isso implica que não existe apenas uma língua de sinais universal. Cada nação possui sua própria língua de sinais, podendo até mesmo existir variações entre regiões e entre comunidades de deficientes auditivos (regionalismos).

2.1.1 Sinais e seus Parâmetros

Conforme mencionado anteriormente, os sinais são os itens lexicais das línguas de sinais, da mesma forma que os morfemas são os itens lexicais das línguas orais. Os sinais são compostos por fonemas que são as unidades básicas da comunicação baseada em sinais. Segundo Buttussi et al. (2007), um sinal consiste e é unicamente identificado por cinco fonemas:

1. **Configuração da mão:** representa a posição dos dedos e seus movimentos. Um sinal pode ser produzido pela mão predominante (mão direita para os destros) ou pelas duas mãos. Uma configuração de mão pode se diferenciar das demais pela extensão (lugar e número de dedos estendidos), pela contração (mão aberta ou fechada), e pelo contato ou divergência dos dedos. Segundo Felipe (2007), em LIBRAS existem atualmente 60 possíveis configurações de mão, dentre as quais 26 são usadas para representar as letras do alfabeto. Na Figura 2.1 são ilustradas as configurações de mão de LIBRAS segundo Felipe (2007). Os sinais APRENDER, LARANJA e ADORAR, por exemplo, possuem a mesma configuração de mão.
2. **Ponto de Articulação:** representa a parte do corpo do emissor onde os sinais são realizados (ou começam a ser realizados). Em LIBRAS, esse ponto de articulação pode estar localizado em alguma parte do corpo ou estar localizado em um espaço neutro vertical (do meio do corpo até a cabeça) e horizontal (à frente do emissor). Por exemplo, os sinais TRABALHAR, BRINCAR, CONSERTAR são executados no espaço neutro, enquanto que os sinais ESQUECER, APRENDER e PENSAR são feitos na testa.

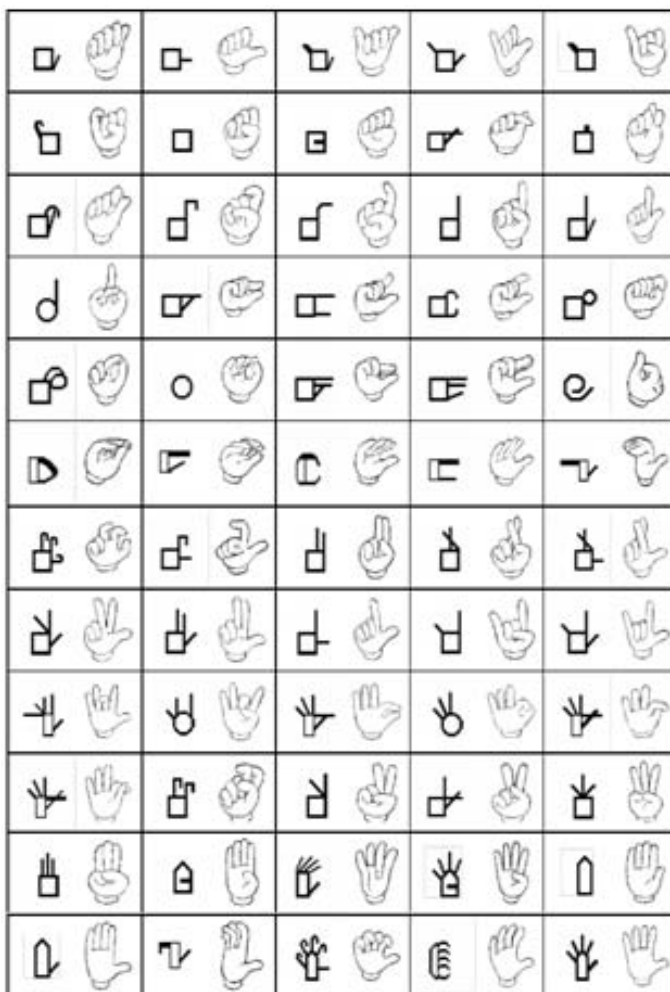
3. **Movimento:** um sinal pode ter movimentos ou não. O movimento representa o deslocamento da mão no espaço durante a execução do sinal. Caracterizar os movimentos não é uma tarefa trivial, uma vez que a mão é um objeto assimétrico e que os eixos podem se deslocar simultaneamente. Segundo Fusco (2004), no entanto, a maioria dos sinais pode ser dividida em pequenos segmentos de movimentos e cada um dos quais relacionados a um eixo. Em LIBRAS, por exemplo, os sinais AJOELHAR e EM-PÉ não tem movimento.
4. **Orientação:** representa a direção ou orientação do movimento. Os sinais que possuem algum tipo de movimento são executados em uma determinada direção. Em LIBRAS, a inversão da direção pode representar a idéia de oposto, contrário ou modificar a concordância número-pessoal do sinal.
5. **Expressões não manuais:** muitos sinais podem requerer características adicionais para expressar sentimentos de alegria, de tristeza, uma pergunta ou uma exclamação e necessitam de um traço diferenciador, como uma expressão facial e/ou corporal, para poder expressar essas características. Dessa forma, as expressões não manuais que podem ser expressões faciais e/ou corporais podem assumir tanto uma função léxica (expressar alegria e tristeza, por exemplo) quanto uma função sintática (expressar uma pergunta ou uma exclamação, por exemplo) na estrutura dos sinais. A inclusão de uma expressão facial e/ou corporal no discurso pode alterar o significado de um sinal e sua ausência pode produzir um sinal sem sentido. Alguns sinais em LIBRAS, inclusive, são feitos apenas com a bochecha, como LADRÃO e ATO-SEXUAL.

Conforme Buttussi et al. (2007), a partir da combinação desses cinco fonemas formam-se os sinais, que permitem comunicar diferentes idéias ou aspectos de uma sentença. Expressar-se em LIBRAS ou em qualquer outra língua de sinais, portanto, é combinar esses elementos para formar os sinais e combinar os sinais para formar sentenças em um determinado contexto.

2.1.2 Espaço de Sinalização

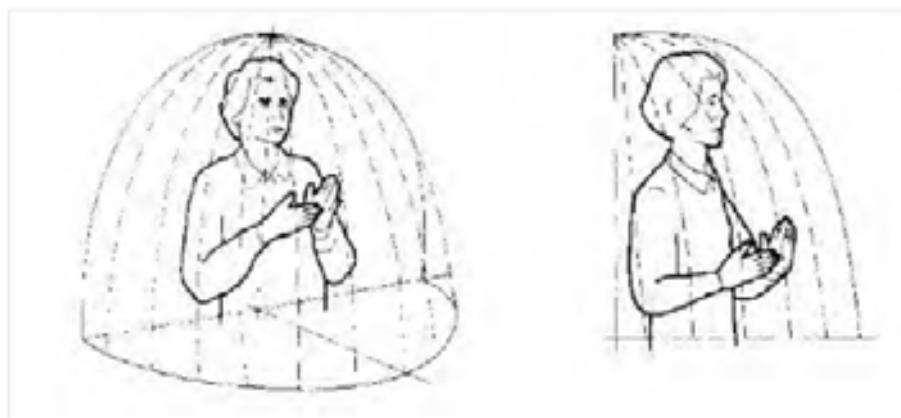
O espaço de sinalização corresponde à área em que os sinais são articulados nas línguas de sinais. Essa área é um espaço tridimensional (3D) que se estende da cabeça até a cintura e se estende para fora do corpo até que o braço esteja completamente estendido. Os movimentos de mão e expressões faciais ou corporais de um sinal ocorrem sempre nesse espaço. Um diagrama desse espaço, extraído de Morrissey (2008), é apresentado na Figura 2.2.

Figura 2.1: Configurações de mão da LIBRAS



Na comunicação em línguas de sinais, esse espaço de sinalização pode ser dividido nos pontos de referência dêitica e no espaço de sinalização neutro. Os pontos de referência dêitica podem ser usados por um sinalizador (isto é, uma pessoa que está representando os sinais) para associar pessoas, objetos ou conceitos em discussão com locais (ou pontos de referência) no espaço de sinalização. Por exemplo, suponha uma cena em que exista uma casa no lado esquerdo da cena. Nesse caso, um sinalizador pode representar o sinal "casa" e apontar para o lado esquerdo no espaço de sinalização. Isso indica que um ponto de referência para o sinal "casa" foi criado no espaço de sinalização. Se o sinalizador quiser representar um homem andando em direção a casa, ele pode simplesmente utilizar esse ponto de referência para indicar que a casa é o ponto final do movimento, isto é, ele não precisa representar o sinal "casa" novamente e pode utilizar seu ponto de referência. A utilização desses pontos de referência no espaço 3D não ocorre nas línguas orais e

Figura 2.2: Diagrama do espaço de sinalização



representa uma importante informação semântica para as sentenças nas línguas de sinais. O espaço neutro, por outro lado, representa todos os outros locais no espaço de sinalização que não contém pontos de referência associados.

2.1.3 Sistema de Representação e Escrita

Uma das diferenças marcantes entre as línguas de sinais e as línguas orais é a nítida falta de uma adoção formal, ou mesmo reconhecida, de um sistema de escrita para as línguas de sinais. Segundo Morrissey (2008), uma das razões é devido a estrutura paralela das línguas de sinais. Conforme mencionado anteriormente, as línguas de sinais possuem uma estrutura paralela, podendo emitir sinais que envolvem simultaneamente diversas partes do corpo do sinalizador, o que dificulta a criação de um sistema de escrita linear.

Existem várias tentativas para se criar sistemas de escrita para línguas de sinais, mas a maioria delas não é utilizada pelo público geral [Morrissey 2008]. Nessa subseção discutiremos os sistemas de escrita e transcrição mais utilizados, como, por exemplo, a notação de Stokoe (1980), o HamNoSys [Hanke 2004] e o SignWriting [Sutton 1995].

A notação de Stokoe (1980) foi desenvolvida na década de 1960 para descrever sinais em ASL, com base em três fatores: *tabulation*, para representar o ponto de articulação do sinal; *designator*, para representar a configuração da mão; e *signation*, para representar o tipo de movimento. Essa notação não possui símbolos para representar expressões faciais e/ou corporais e não é prática para ser utilizada pelos deficientes auditivos. Na Figura 2.3, extraída de Morrissey (2008), é apresentado o sinal "don't know" na notação de Stokoe.



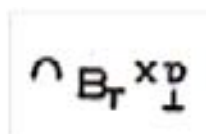
O primeiro símbolo  é um símbolo de ponto de articulação que representa a parte superior da testa. O símbolo  representa uma configuração de mão, onde todos os

Figura 2.3: Representação do sinal "don't know" na notação de Stokoe



dedos estão estendidos e lado a lado, mas com o polegar dobrado e com a palma da mão direcionada para o sinalizador. O símbolo $\overset{x}{\cdot}$ representa o movimento da mão para tocar a testa e o símbolo $\overset{p}{\perp}$ representa a palma da mão virada para baixo e um movimento com a mão se afastando do sinalizador.

Outra notação escrita para línguas de sinais é a *Hamburg Notation System* (HamNoSys) [Hanke 2004]. Essa notação utiliza um conjunto de símbolos independentes para representar iconicamente os fonemas das línguas de sinais. Um dos problemas da notação HamNoSys é que ela, assim como a notação de Stokoe, não possui símbolos para representar expressões faciais e/ou corporais. Na Figura 2.4, extraída de Morrissey (2008), é apresentado o sinal "nineteen" na notação HamNoSys.

Figura 2.4: Representação do sinal "nineteen" na notação HamNoSys



A notação SignWriting, desenvolvida por Sutton (1995), utiliza símbolos intuitivos e icônicos para expressar visualmente os movimentos de mão, os pontos de articulação e também expressões faciais e/ou corporais. Essa notação vem sendo aprendida por crianças e adultos em todo o mundo, como uma versão escrita das línguas de sinais. Ela ainda não é uma notação largamente utilizada, mas o número de livros publicados em SignWriting vem crescendo nos últimos anos [Morrissey 2008]. Na Figura 2.5, também extraída de Morrissey (2008), é apresentado o sinal "deaf" em SignWriting.

Figura 2.5: Representação do sinal "deaf" na notação SignWriting



Por fim, na representação em glosa, os sinais são representados textualmente utili-

zando os mesmos símbolos e a mesma representação da língua oral. Essa estratégia não resolve o problema do sistema de escrita, mas facilita o processo de representação das línguas de sinais. Por exemplo, pode-se utilizar a representação "GATO" para o sinal correspondente ao animal gato. As sentenças, no entanto, são estruturadas de acordo com a estrutura gramatical da língua de sinais.

2.1.4 LIBRAS

Conforme mencionado anteriormente, LIBRAS é a língua de sinais oficial do Brasil. De forma similar as outras línguas de sinais, os sinais de LIBRAS são constituídos e identificados pelos cinco fonemas (configuração de mão, ponto de articulação, movimento, direção e expressão facial e/ou corporal) apresentados na Seção 2.1.1. Os possíveis valores para cada um desses fonemas são discutidos e detalhados em Fusco (2004). Os possíveis valores para configuração de mão, em especial, são apresentadas na Figura 2.1.

Em LIBRAS, existem algumas restrições fonológicas que servem para ajudar na composição dos sinais. Segundo Quadros (1997), os sinais podem ser produzidos utilizando uma das mãos ou as duas. Quando as duas mãos são utilizadas, é possível ter as duas mãos ativas na representação do sinal ou a mão secundária funcionar apenas como localização (ponto de articulação). Outro tipo de restrição diz respeito à simetria e a dominância de uma das mãos. No primeiro caso, tem-se a mesma configuração de mão e o mesmo ponto de articulação para as duas mãos. No segundo caso, as mãos podem ter diferentes configurações de mão e enquanto a mão dominante realiza o movimento, a mão secundária funciona apenas como ponto de articulação.

Além dos sinais serem formados pelos mesmos fonemas das demais línguas de sinais, a LIBRAS também possui uma gramática própria, diferente da gramática da língua portuguesa. Com relação à ordem das palavras ou constituintes, por exemplo, existem diferenças entre a língua portuguesa e a LIBRAS. Enquanto que a língua portuguesa utiliza, na maioria dos casos, sentenças no formato sujeito-verbo-objeto (SVO), a LIBRAS utiliza, em geral, sentenças no formato tópico-comentário [Brito 1995]. Por exemplo, as sentenças em língua portuguesa:

- O urso (S) matou (V) o leão (O).
- Eu (S) não vi (V) o acidente na rua (O).

Seriam representadas em LIBRAS da seguinte forma:

- URSO (Tópico), LEÃO MATAR (Comentário).
- RUA ACIDENTE (Tópico) NÃO ENXERGAR (Comentário).

Embora a estruturação das sentenças em LIBRAS, no que se refere à ordem dos argumentos, seja diferente da estruturação das sentenças em língua portuguesa, existem algumas semelhanças na estrutura das sentenças. Segundo Brito (1995), em ambas as línguas, "toda sentença possui um núcleo que é o elemento que possui valência". Tanto na LIBRAS quanto na língua portuguesa, o verbo é o elemento que possui valência e determina o número e o tipo de argumentos ou complementos necessários. O verbo "enviar", em português, e ENVIAR, em LIBRAS, por exemplo, são verbos com a mesma valência, porque pedem três argumentos. Por exemplo.

- Paulo enviou o livro ao amigo. (em língua portuguesa)
- LIVRO AMIGO P-A-U-L-O ENVIAR. (em LIBRAS)

Conforme pode ser observado, nos dois exemplos, independente da ordem das palavras, as sentenças são constituídas de um núcleo (o verbo enviar) e três argumentos ou complementos (Paulo, amigo e livro). Outra característica que também pode ser observada é que em LIBRAS, os nomes são representados soletrando-se as letras do nome (por exemplo, o nome Paulo é representado em LIBRAS como P-A-U-L-O).

Por fim, a LIBRAS também possui diferenças regionais. Alguns sinais podem ser representados diferentemente de acordo com a região do país. Por exemplo, alguns sinais em LIBRAS pode ser representados diferentemente no nordeste, sul e sudeste do Brasil.

Na próxima seção serão apresentados os principais conceitos relacionados a sistemas de tradução automática.

2.2 Sistemas de Tradução Automática

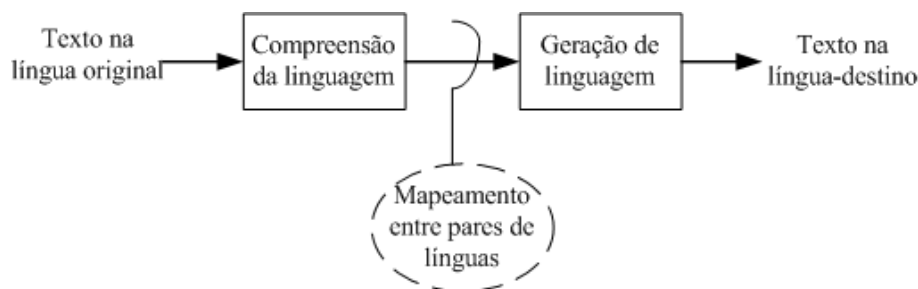
A tradução automática (do inglês, *machine translation* - MT) é a conversão de conteúdos entre línguas naturais através de sistemas computacionais. Esse processo de conversão ou tradução, no entanto, possui uma série de dificuldades e desafios intrínsecos. Por exemplo, quando um mensagem é transmitida de um interlocutor para outro, ela está, em geral, relacionada a um contexto e a um conjunto de conhecimentos de senso comum implícitos que precisam ser capturados e tratados pelo sistema de tradução para se gerar uma boa tradução. Além disso, as línguas naturais, em geral, possuem um conjunto de ambigüidades inerentes (ambigüidade léxica, sintática, semântica, contextual, etc.) que também precisam ser tratadas pelo sistema de tradução automática [Dorr et al. 1999]. De acordo com Dorr et al. (1999), esses desafios podem ser caracterizados em duas classes:

1. Desafios com relação aos aspectos lingüísticos

2. Desafios com relação aos aspectos operacionais

Com relação aos aspectos lingüísticos, os desafios podem ser classificados em três categorias: (1) compreensão de linguagem, (2) geração de linguagem e (3) mapeamento entre pares de linguagens. Esses três tipos de desafios estão relacionados, conforme ilustrado na Figura 2.6, adaptada de [Dorr et al. 1999].

Figura 2.6: Três tipos de desafios lingüísticos dos sistemas de tradução automática.



Os desafios relacionados a compreensão de linguagem são aqueles desafios relacionados a compreensão do texto de entrada na língua original. Dentre esses desafios, destacam-se a ambigüidade léxica, sintática, semântica e contextual. Exemplos de cada um desses tipos de ambigüidade são apresentados em [Dorr et al. 1999]. Os desafios relacionados a geração de linguagem são aqueles que estão relacionados com a geração de textos na língua alvo. Um dos principais desafios nesse tipo de atividade é a seleção léxica, que corresponde a dificuldade de selecionar adequadamente as palavras na língua-destino que representam o conhecimento que está embutido na sentença na língua-original. Por exemplo, identificar se a palavra "esperar", em português, seria melhor traduzida como "hope" ou "wait" em inglês. Por fim, os desafios relacionados ao mapeamento entre pares de linguagens estão relacionados as diferenças gramaticais entre línguas. Por exemplo, variação de posicionamento entre sujeito e objeto que ocorre em diversas línguas [Dorr et al. 1999].

Com relação aos aspectos operacionais, os principais desafios estão relacionados a dificuldade de estender os sistemas de tradução para novos domínios ou línguas; as dificuldades de manter o sistema após o seu desenvolvimento, as dificuldades de integração com outras ferramentas e as métricas para avaliar a qualidade das traduções.

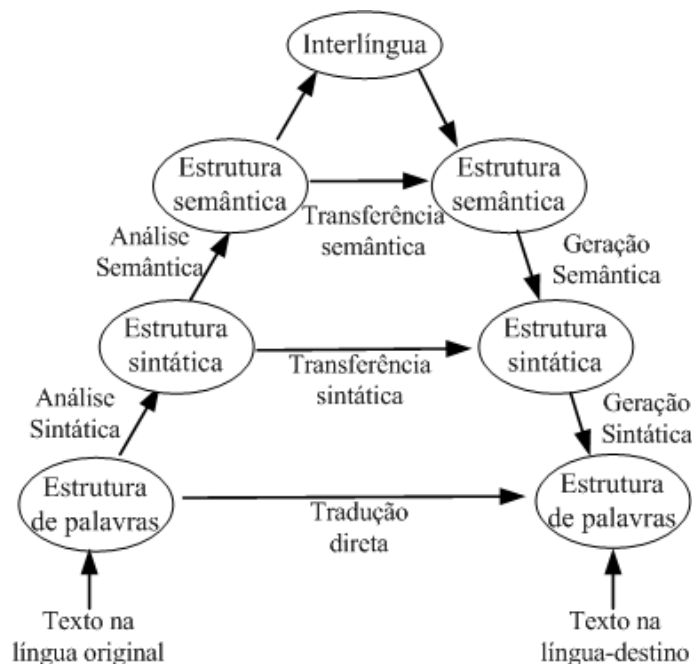
Em virtude desses desafios, diversos sistemas e diferentes abordagens de tradução automática vêm sendo desenvolvidas nos últimos anos [Dorr et al. 1999][Su & Wu 2009]. De acordo com Su & Wu (2009), eles podem ser divididos em três classes principais: Tradução Automática Baseada em Regras (*Rule-Based Machine Translation* - RBMT),

Tradução Automática Estatística (*Statistical Machine Translation* - SMT) e Tradução Automática Baseada em Exemplos (*Example-Based Machine Translation* - EBMT). Nas próximas seções, cada uma dessas abordagens será apresentada.

2.2.1 Tradução Automática Baseada em Regras

Os sistemas de Tradução Automática Baseada em Regras (RBMT) traduzem um texto de uma língua original para uma língua alvo usando informações lingüísticas desenvolvidas manualmente por especialistas proficientes nas duas línguas. De acordo com Morrissey (2008), esses sistemas podem ser divididos em três tipos básicos: (1) sistemas de tradução direta, (2) sistemas de tradução baseada em transferência e (3) sistema de tradução baseado em interlíngua. Esses tipos de sistemas são usualmente ilustrados em um diagrama piramidal, conforme pode ser observado na Figura 2.7, adaptada de Morrissey (2008).

Figura 2.7: Tipos de sistemas de tradução automática baseado em regras



Os sistemas de tradução direta aplicam uma tradução palavra-por-palavra da língua original para a língua-destino, isto é, sem realizar qualquer análise sintática ou semântica. O resultado deste tipo de tradução é um texto (na língua-destino) formado a partir da substituição direta das palavras na língua-original. Isso implica que a ordem das palavras no texto traduzido será a mesma ordem das palavras na língua-original, mesmo que a

língua-destino não permita esse tipo de estruturação (ordenação das palavras). Isso implica que o leitor, em geral, precisa conhecer a estruturação sintática da língua-original para compreender o texto transmitido.

Outro problema enfrentado por esse tipo de sistema é a escolha da tradução correta para cada palavra na língua-original, denominado de ambigüidade léxica. Por exemplo, a palavra "book", em inglês, pode ser traduzida para "livro" ou "reservar", dependendo do contexto. Dessa forma, os textos traduzidos utilizando esse tipo de arquitetura são, em geral, difíceis de compreender. No entanto, para traduções em textos simples que possuam domínios bem restritos, eles podem, em algumas situações, produzir textos úteis.

Os sistemas de tradução baseada em transferência utilizam estruturas gramaticais com regras de tradução (ou exemplos) nos níveis sintático ou semântico para prover traduções mais complexas que as providas pelos sistemas de tradução direta. Mais especificamente, nesse tipo de sistema, o texto na língua-original é primeiramente convertido em uma representação sintática (ou semântica) interna na língua-original (Análise Sintática ou Semântica) com base na representação gramatical utilizada. Com base nas regras de tradução (ou exemplos), esta representação é então convertida em uma representação sintática (ou semântica) interna na língua-destino. As regras de tradução, portanto, mapeiam uma representação sintática (ou semântica) da língua-original em uma representação equivalente na língua-destino. Essa representação na língua-destino é, então, convertida para uma representação textual nessa língua (Geração Sintática ou Semântica). Por exemplo, nos sistemas de tradução baseado em transferência para línguas de sinais, estruturas gramaticais como *Lexical Functional Grammar* ou *Synchronous Tree Adjoining Grammar* [va Zijl & Olivrin 2008][Veale et al. 1998] são utilizadas.

Uma das vantagens das abordagens baseadas em transferência é que elas são capazes de resolver algumas ambigüidades, como, por exemplo, as ambigüidades léxicas. Uma vez que a análise sintática, em geral, permite identificar a categoria léxica das palavras no texto-original, é possível determinar, por exemplo, que a palavra "book", em inglês, seria traduzida para o substantivo "livro" ou para o verbo "reservar", dependendo do contexto local. No entanto, algumas ambigüidades mais complicadas não são resolvidas, a menos que exista uma similaridade entre as línguas [Dorr et al. 1999], como ocorre na relação entre a língua oral e a língua de sinais de uma país (como, por exemplo, ente Língua Portuguesa e LIBRAS) [Woodward 1973].

A adição de informações semânticas (sistemas de tradução baseada em transferência semântica), no entanto, auxilia na resolução de algumas dessas ambigüidades. Por exemplo, considere a seguinte sentença: "João golpeou o cachorro com uma vara". Nesse exemplo, a análise sintática não é capaz de identificar se "o cachorro estava com a vara" ou

se "João usou a vara para golpear o cachorro". No entanto, com o auxílio de informações semânticas (ou de contexto), é possível solucionar esta ambigüidade. Por exemplo, se houvesse alguma informação semântica anterior no texto que identificasse que "João estava com uma vara e estava sendo ameaçado pelo cachorro".

Em resumo, os sistemas de tradução baseados em transferência produzem traduções melhores que os sistemas de tradução direta, mas tem a necessidade de desenvolver estratégias para análise sintática ou semântica, além de ter que desenvolver um conjunto de regras de tradução (ou exemplos).

Por fim, os sistemas de tradução baseados em interlíngua utilizam uma linguagem intermediária de representação, independente da língua-original, denominada interlíngua. Nesse tipo de sistema de tradução, o texto original é analisado e semanticamente processado, gerando uma representação na interlíngua. Em seguida, a representação na interlíngua é convertida para a representação do texto na língua-destino. Uma vez que uma interlíngua, independente da língua-original, é utilizada, é possível incluir novas línguas no sistemas apenas adicionando uma etapa de análise e geração para a nova língua. Conforme comentado anteriormente, em um sistema de tradução baseado em transferência, além das etapas de análise e geração, também seria necessário definir um novo conjunto de regras de tradução para cada língua do sistema.

Um dos problemas encontrados neste tipo de abordagem é que como as representações em interlíngua são independentes da sintaxe da língua-original, a geração dos textos na língua-destino tendem a perder o estilo e ênfase do texto original. Contudo, excluindo os textos artísticos (ficção ou poesias), a preservação do estilo de texto do autor é, em geral, supérfluo [Dorr et al. 1999].

2.2.2 Tradução Automática Estatística

Diferentemente das estratégias de RBMT apresentadas na Seção 2.2.1, a Tradução Automática Estatística (SMT) não é baseada em uma representação sintática ou semântica formal nas duas línguas. Ao invés disso, esses sistemas se baseiam na tradução mais provável com base em um corpus paralelo bilingüe¹. Em outras palavras, a tradução automática estatística procura extrair automaticamente o conhecimento (regras de tradução e base de conhecimento) acerca da tradução a partir desse corpus.

Considere, por exemplo, o problema de traduzir uma sentença E , em inglês, para uma sentença F em francês. Na tradução automática estatística, pode-se expressar esse

¹Um corpus paralelo bilingüe é uma coleção de textos em uma língua L_1 e sua correspondente tradução em uma língua L_2

problema de tradução, pela aplicação da Regra de Bayes a seguir:

$$\operatorname{argmax}_F(P(F|E)) = \operatorname{argmax}_F \frac{(P(E|F)P(F))}{P(E)}, \quad (2.1)$$

$$\operatorname{argmax}_F(P(F|E)) = \operatorname{argmax}_F(P(E|F)P(F)), \quad (2.2)$$

Onde, $P(F|E)$ representa a probabilidade da sentença E ser traduzida como a sentença F , de acordo com o corpus bilingüe; $P(E|F)$ representa a probabilidade de F ser traduzida como E ; $P(E)$ e $P(F)$ representam a probabilidade das sentenças E e F no corpus da língua inglesa e francesa, respectivamente.

Essa regra indica que, considerando todas as sentenças possíveis em francês F , deve-se escolher aquela que maximiza o produto $P(E|F)P(F)$. O fator $P(E)$ pode ser ignorado, pois ele possui o mesmo valor para todas as sentenças F . O fator $P(F)$ é denominado o **modelo da linguagem** para o francês, e informa qual a probabilidade de uma dada sentença estar em francês. $P(E|F)$ é denominado **modelo de tradução**, e informa qual a probabilidade de uma sentença em inglês ser uma tradução de uma dada sentença em francês.

Analisando a equação acima, observa-se que ela não possui uma analogia clara com o processo utilizado por um tradutor humano. Não é intuitivo imaginar um tradutor humano, mentalmente, analisando a lista de todas as sentenças em francês para calcular o produto da probabilidade $P(F)$ com a probabilidade condicional, $P(E|F)$, de tradução da sentença em inglês para cada sentença em francês.

A justificativa para se aplicar a regra de Bayes nesse caso é, em essência, para tornar possível a divisão do problema em duas partes. Primeiro, utiliza-se o modelo de tradução $P(E|F)$ para encontrar sentenças candidatas em francês. Algumas dessas sentenças, no entanto, podem não estar em francês fluente. Em seguida, utiliza-se o modelo de linguagem $P(F)$, para estimar qual das sentenças candidatas é provavelmente a tradução mais comum e mais fluente.

O modelo de linguagem, $P(F)$, pode ser qualquer modelo que forneça uma probabilidade para uma dada sentença F . Uma alternativa para implementar esse modelo seria estimar $P(F)$ diretamente, contando quantas vezes cada sentença aparece no corpus. Contudo, mesmo com um corpus muito grande, a maioria das sentenças nunca aparecerá no corpus, e dessa forma, a sua probabilidade será zero. Para solucionar esse problema, uma alternativa é utilizar o modelo de linguagem de bigramas, em que a probabilidade de uma sentença F em francês composta pelas palavras f_1, \dots, f_n é definida da seguinte forma:

$$P(f_1, \dots, f_n) = \prod_{i=1}^n P(f_i | f_{(i-1)}) \quad (2.3)$$

Essa estratégia, no entanto, é capaz de captar apenas uma noção de sintaxe muito local, onde uma palavra depende apenas da palavra anterior. Por exemplo, seria necessário conhecer as probabilidades de bigramas como $P(\text{"torre Eiffell"})$. Para tradução direta, no entanto, isso é, em geral, suficiente. A opção mais comum, no entanto, é utilizar modelos de trigramas, em que uma palavra depende das duas palavras anteriores.

O modelo de tradução, $P(E|F)$, possui uma complexidade maior, uma vez que ele considera o produto de pares de sentenças (inglês, francês), e não apenas sentenças individuais. Um modelo de tradução simplificado seria traduzir cada palavra individual e independentemente, da esquerda para a direita. Esse modelo computaria a probabilidade de uma tradução da seguinte forma:

$$P(E|F) = \prod_{i=1}^n P(E_i|F_i) \quad (2.4)$$

Um dos problemas desse modelo é a ordem das palavras. Uma sentença em inglês da forma [Adjetivo + Substantivo], por exemplo, é traduzida para o francês na ordem [Substantivo + Adjetivo]. Além disso, conforme comentado anteriormente, a escolha de palavras não é necessariamente um mapeamento de um para um (ambigüidade léxica).

Um modelo de tradução um pouco mais sofisticado é o modelo **IBM Model 3** [Brown et al. 1993] que continua utilizando unigramas, isto é, tradução palavra por palavra, mas acrescenta algumas características para lidar com os problemas apresentados. Uma das características introduzidas por esse modelo é o conceito de fertilidade de uma palavra. A fertilidade representa o número de palavras que uma palavra gera na outra língua. Uma palavra com fertilidade n é copiada n vezes e cada uma dessas cópias é traduzida independentemente. Por exemplo, para traduzir a expressão "*à la maison*" do francês para "*home*" em inglês, o modelo definiria uma fertilidade 0 para as palavras "*à*" e "*la*" e uma fertilidade 1 para a palavra "*maison*" e depois utilizaria esse modelo para traduzir "*maison*" como "*home*". Para traduções no outro sentido (inglês para francês), a palavra "*home*" teria fertilidade 3 e seria copiada 3 vezes, gerando "*home home home*". Em seguida, a primeira palavra "*home*" seria traduzida para "*à*", a segunda seria traduzida para "*la*" e a terceira como "*maison*".

Para lidar com o problema da ordem das palavras, o modelo permuta as palavras na posição correta, através de deslocamentos que movem a palavra de sua posição original para a posição final. Essa operação é, em geral, denominada alinhamento das palavras.

No **IBM Model 3**, o deslocamento leva em consideração a posição da palavra dentro da sentença e o comprimento da sentença em ambos os idiomas. Isto é, o modelo avalia os seguintes parâmetros:

$$P(\text{Desloc} = o | \text{Pos} = p, \text{EngLen} = m, \text{FrLen} = n), \quad (2.5)$$

Onde, *Pos* corresponde a posição da palavra na sentença, *EngLen* representa o comprimento da sentença em inglês e *FrLen* representa o comprimento da sentença em francês.

Por exemplo, as palavras em inglês "dog" e "brown" são traduzidas em francês como "chien" e "brun", respectivamente, mas a tradução de "brown dog" é "chien brun". Portanto, poderia-se determinar o deslocamento de "brown" em "brown dog", consultando $P(\text{Desloc}|1,2,2)$. Essa modelo então poderia indicar, por exemplo, um deslocamento +1 (ela é movida uma posição para a direita) com probabilidade 0,7 maior do que um deslocamento 0 (ela não é movida), por exemplo, com probabilidade 0,3. Analogamente, o deslocamento da palavra "dog", $P(\text{Desloc}|2,2,2)$, deveria indicar um deslocamento -1 (ela é movida uma posição para a esquerda) com probabilidade maior do que um deslocamento 0.

2.2.3 Tradução Automática Baseada em Exemplos

A Tradução Automática Baseada em Exemplos (EBMT) proposta por Nagao (1984), é uma estratégia de "tradução por analogia", que procura similaridades em traduções anteriores para construir traduções para novas sentenças. De acordo com Nagao (1984), a inspiração da EBMT é o fato de que os seres humanos não fazem traduções utilizando análises lingüísticas profundas das sentenças. Em vez disso, os seres humanos traduzem uma sentença, decompondo-a em um conjunto de fragmentos, traduzindo esses fragmentos a partir de analogias com traduções anteriores (exemplos) e combinando esses fragmentos traduzidos para formar a sentença traduzida.

Em virtude da EBMT ser baseada em analogias de traduções anteriores, os sistemas EBMTs, assim como os sistemas SMT, também necessitam de corpus paralelos bilingües (conjunto de exemplos) para realizar suas traduções. O processo de tradução de um sistema EBMT inicia com uma etapa de aquisição de exemplos a partir do corpus bilingüe [Kit et al. 2002]. Nessa etapa, os exemplos são extraídos do corpus bilingüe através de um algoritmo de alinhamento de textos. Esse algoritmo é responsável por identificar a correspondência (ou o casamento) entre pares de sentenças entre os dois corpus (na língua original e na língua de destino) e essa correspondência entre as sentenças é, em geral, definida com base em uma medida de similaridade. Em outras palavras, o alinhamento

de texto identifica para cada sentença do corpus na língua original qual é a sentença (tradução) correspondente no corpus na língua de destino. Uma outra alternativa é fazer o alinhamento manual dos textos, onde especialistas determinam a correspondência entre as sentenças, mas, em geral, o uso de alinhamento manual esbarra na dificuldade de alinhar corpus muito extensos (com milhões ou dezenas de milhões de palavras). Os exemplos coletados nessa etapa de alinhamento são então organizados e armazenados numa "base de exemplos"(BE).

Após a criação da BE, uma etapa de aplicação dos exemplos é aplicada para traduzir a sentença de entrada [Kit et al. 2002]. Nessa etapa, a sentença de entrada é, inicialmente, decomposta em uma seqüência de fragmentos de sentenças. A partir desses fragmentos, o sistema EBMT, então, pesquisa quais são os exemplos na BE que mais se aproximam dos fragmentos resultantes do processo de decomposição. A proximidade entre esses fragmentos é, em geral, definida utilizando "medidas de distância" entre as sentenças, baseadas, por exemplo, em tesouros² ou ontologias³ [Dorr et al. 1999]. Com base nos exemplos mais similares (analogia), esses fragmentos de sentença são então traduzidos pelo EBMT da língua original para a língua alvo.

Por fim, uma etapa de síntese da sentença alvo é aplicada. Essa etapa consiste em combinar os fragmentos traduzidos para formar uma sentença legível na língua de destino [Kit et al. 2002]. Formar uma sentença legível na língua de destino a partir de um conjunto de fragmentos, no entanto, não é uma tarefa trivial. Essa etapa é considerada, inclusive, a mais difícil de ser desenvolvida num processo tradução automática baseada em exemplos [Kit et al. 2002].

Uma característica importante dos sistemas EMBT, é que a precisão e qualidade da tradução dependem do tamanho e da cobertura do corpus paralelo bilingüe, que precisa ser, em geral, bastante extenso para cobrir as diferenças sintáticas e semânticas entre as línguas envolvidas. Em virtude disso e das dificuldades para sintetizar as sentenças, a maioria dos sistemas EBMT são restritos a partes das sentenças, como, por exemplo, partes preposicionais ou nominais [Sato 1993], dentre outras.

²Os tesouros, também conhecido como dicionário de ideias afins, são uma lista de palavras com significados semelhantes, dentro de um domínio específico de conhecimento.

³Em Ciência da Computação, uma ontologia é um modelo de dados que representa um conjunto de conceitos dentro de um domínio e os relacionamentos entre estes conceitos

2.3 Computação Humana

Embora os sistemas computacionais venham superando os seres humanos em diversas tarefas, especialmente tarefas que são tediosas ou que consomem muito tempo, eles ainda possuem grandes dificuldades em algumas tarefas que são naturais para os humanos, como, por exemplo, identificar uma pessoa em uma cena (processamento visual) [Yuen et al. 2009]. Outras exemplos de atividades que os computadores não conseguem ou não resolvem bem sozinhos são o processamento de linguagem naturais, a coleta de conhecimentos de senso comum, o raciocínio, dentre outros.

Uma alternativa para solucionar esses problemas é enxergar ou utilizar os seres humanos como unidades de processamento (processadores), integrando-os no processo de resolução do problema. Esse princípio é o que norteia o paradigma da computação humana (*human computation*).

De acordo com Quinn & Bederson (2011), a computação humana (*human computation*) é um paradigma que utiliza o poder de processamento humano para resolver os problemas que os computadores ainda não são capazes de resolver sozinhos. Mais especificamente, a sua proposta é que os processos computacionais deleguem para os seres humanos a execução de algumas etapas do problema que eles ainda não são capazes de resolver, tratando as pessoas como processadores capazes de executar tarefas pontuais que resolvem partes de problemas complexos de computação massiva [Ahn & Dabbish 2008].

Para viabilizar a participação das pessoas, no entanto, é preciso motivá-las [Quinn & Bederson 2011]. Essa motivação pode ser financeira, remunerando as pessoas por sua colaboração ou recompensando-as por bons resultados, por exemplo. Outra forma de motivação é indireta. Nesse tipo de motivação, a colaboração acontece como efeito colateral de alguma atividade realizada, como por exemplo, através de jogos. As pessoas participam da computação não para resolver um problema específico, mas para se entreterem. A solução do problema surge como um efeito colateral resultado da computação executada durante o jogo. Em consequência disso, os sistemas de computação humana que utilizam esse tipo de abordagem são denominados de Jogos com Propósito (do inglês, *Game With A Purpose* - GWAP) [Ahn & Dabbish 2008].

Além da motivação, de acordo com Quinn & Bederson (2011), existem também outros fatores que caracterizam os sistemas de computação humana. São eles:

- Controle de qualidade das contribuições;
- Estratégia de agregação das contribuições;
- Habilidades humanas requisitadas pelo sistema;
- Ordem do processo;

- Cardinalidade na relação entre tarefas e trabalhadores (participantes).

O controle de qualidade das contribuições de um sistema de computação humana está relacionado as estratégias utilizadas para garantir a qualidade das informações produzidas pelos participantes. O uso de estratégias é importante nesse tipo de sistema, porque mesmo que os participantes estejam motivados e não tenham intenção de enganar ou sabotar o sistema, eles podem não compreender as direções do sistema ou mesmo cometer erros. Dentre as estratégias de controle de qualidade usualmente usadas, pode-se destacar a concordância na saída, em que a resposta é aceita apenas se duas pessoas concordarem com a mesma resposta, a filtragem estatística, que consistem em eliminar os *outliers* (ou pontos fora da curva estatística), a revisão (ou supervisão) das respostas por especialistas, dentre outros [Quinn & Bederson 2011].

A agregação das contribuições está relacionada a estratégia usada para combinar as contribuições individuais e resolver o problema global. Dentre as principais estratégias utilizadas nos sistemas de computação humana pode-se destacar a coleção, em que uma base de conhecimento de dados discretos (independentes) é coletada, o processamento estatístico dos dados (*the wisdom of crowds*⁴, por exemplo), dentre outros.

As habilidades humanas requisitadas pelo sistema, como o próprio nome indica, representam as habilidades gerais ou específicas dos participantes (trabalhadores) necessárias para solucionar as tarefas. Em um sistema de computação humana para tradução para chinês, por exemplo, o participante precisa dominar a língua chinesa.

A ordem do processo está relacionada a ordem em que os envolvidos no sistema de computação humana (solicitantes, trabalhadores e computador) atuam no sistema. Por exemplo, o sistema está na ordem "Computador + Trabalhador + Solicitante"[Ahn et al. 2008] quando o computador inicialmente processa alguma informação. Em seguida, os humanos (trabalhadores) auxiliam nesse processamento e a informação resultante é utilizada pelo solicitante (usuário final que se beneficia da computação). Outras ordenações, no entanto, também são possíveis. Os sistemas podem estar na ordem "Trabalhador + Solicitante + Computador"[Ahn & Dabbish 2004], na ordem "Computador + Trabalhador + Solicitante + Computador"[FACTory 2012], dentre outras.

Por fim, a cardinalidade está relacionada com a quantidade de trabalhadores e de tarefas envolvidos no sistema. Essa cardinalidade está, em geral, relacionada a estrutura do problema. Por exemplo, o sistema pode ser de "um para um", quando um único usuário

⁴A sabedoria das multidões *the wisdom of crowds* é um conceito formulado por Surowiecki (2005) que afirma que a agregação de respostas a partir de um grupo desorganizado e descentralizado de pessoas com pensamentos independentes pode render resultados precisos a perguntas que seriam difíceis de serem respondidas por uma única pessoa

(trabalhador) realiza o processamento de uma única tarefa; "muitos para muitos", quando vários trabalhadores estão relacionados a diversas tarefas; "muitos para um"; dentre outros.

Nas próximas subseções serão apresentados alguns exemplos de sistemas de computação humana, como, por exemplo, os jogos com propósito e o projeto reCAPTCHA.

2.3.1 Jogos com Propósito

De acordo com uma pesquisa realizada pela *International Game Design Association* (IGDA), cerca de 200 milhões de pessoas participam de jogos *online* por semana [IGDA 2012]. Com o objetivo de encorajar esse elevado número de usuários potenciais a prover informações para resolver problemas computacionalmente intratáveis, os Jogos com Propósito (GWAPs) são uma estratégia de computação humana que procura gerar informações como efeito colateral do entretenimento desses usuários quando participam do jogo. Essas informações podem ser usadas para solucionar problemas computacionais diretamente ou para treinar algoritmos de inteligência artificial [Yuen et al. 2009].

Uma característica importante dos Jogos com Propósito é que eles não dependem de altruísmo ou de incentivos financeiros dados às pessoas para funcionar. Ele é, em geral, um jogo que as pessoas querem jogar, e a computação útil é extraída como um efeito colateral. Além disso, suas regras devem incentivar os jogadores a seguir corretamente os passos para resolver o problema computacional e, se possível, devem maximizar a probabilidade da saída obtida ser correta.

Segundo Ahn & Dabbish (2008) existem três tipos de Jogos com Propósito: (1) os jogos em que os usuários recebem a mesma entrada e devem encontrar a mesma saída, de forma independente, (2) os jogos de inversão, em que um usuário recebe uma entrada e o outro deve descobrir qual é a entrada e (3) jogos em os que usuários determinam se receberam a mesma entrada.

Como exemplos de Jogos com Propósito podemos citar o ESP [Ahn & Dabbish 2004] e o Peekaboom [Ahn et al. 2006]. O ESP é um jogo que procura tornar a tediosa e custosa tarefa de rotular imagens manualmente em um jogo onde a saída é um conjunto de rótulos para diferentes imagens disponíveis na internet. De forma resumida, o ESP é um jogo cooperativo para duas pessoas, onde ambos recebem uma imagem e devem descobrir o que seu parceiro digitou com relação àquela imagem (a comunicação entre os jogadores é proibida). Quando os dois jogadores fornecem a mesma saída para uma dada imagem, os dois recebem a mesma quantidade de pontos e partem para a próxima imagem. O Peekaboom, por outro lado, é um jogo *online* no qual as pessoas, ao jogarem,

colaboram com a construção de uma base de dados para treinamento de algoritmos de visão computacional [Ahn et al. 2006].

2.3.2 Projeto reCAPTCHA

O CAPTCHA é um teste amplamente utilizado na Internet para determinar se o usuário do sistema é um humano ou um agente computacional e servem para proteger serviços de internet contra *spammers* [Ahn et al. 2008]. Para isso, ele apresenta um série de caracteres distorcidos que devem ser identificados pelo usuário. Uma vez que os sistemas computacionais não identificam caracteres distorcidos tão bem quanto os seres humanos, é possível identificar se o usuário é humano ou não e, conseqüentemente, proteger o sistema desses *spammers* ou *bots*.

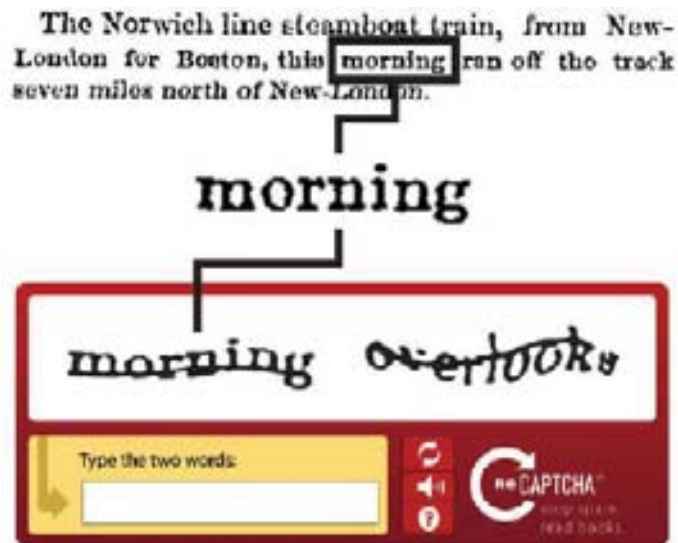
O reCAPTCHA, é um tipo de teste CAPTCHA, que apresenta palavras digitalizadas de textos impressos que não foram capazes de reconhecidas por sistemas de reconhecimento de caracteres (*Optical Character Recognition* - OCR) [Coimbra 2011]. Para diferenciar os humanos dos agentes computacionais, o reCAPTCHA apresenta duas palavras por vez aos usuários: uma palavra de controle conhecida e uma palavra desconhecida para ser reconhecida. Caso o usuário digite corretamente a palavra de controle, o sistema assume que o mesmo é um ser humano e que potencialmente também digitou o significado da outra palavra corretamente. Na Figura 2.8, extraída de [Coimbra 2011], é apresentado um reCAPTCHA com duas palavras: "*morning*" que não foi reconhecida pelo sistema de OCR e é apresentada ao usuário com uma pequena distorção, e "*overlooks*" que funciona como palavra de controle.

Explorando isso, o reCAPTCHA permite que textos sejam digitalizados sem custos. A motivação dos usuários para participar é a vontade de acessar o serviço na Internet que requisita o teste reCAPTCHA.

2.4 Considerações

Nesse capítulo foi apresentada uma fundamentação teórica detalhada sobre línguas de sinais, sistemas de tradução automática e computação humana, os principais conceitos relacionados ao presente trabalho. Na Seção 2.1, foram apresentados os principais conceitos relacionados as línguas de sinais, incluindo o conceito de sinais, espaço de sinalização, sistemas de escrita para línguas de sinais, e as principais características da LIBRAS, a língua de sinais foco deste trabalho. Na Seção 2.2 foram apresentados os principais conceitos relacionados aos sistemas de tradução automática, incluindo as suas

Figura 2.8: Exemplo de reCAPTCHA.



principais características, desafios e abordagens (Tradução Automática Baseada em Regras, Tradução Automática Estatística e Tradução Automática Baseada em Exemplos). Por fim, na Seção 2.3, foram apresentados os principais conceitos relacionados a computação humana, incluindo alguns exemplos de sistemas de computação humana como os jogos com propósito.

No próximo capítulo, será apresentada uma revisão sistemática da literatura direcionada aos sistemas de tradução automática para línguas de sinais, tema de pesquisa desse trabalho. Essa revisão será conduzida para construir um mapa sobre o tema e identificar as principais contribuições e limitações das soluções existentes na literatura científica.

Capítulo 3

Revisão Sistemática da Literatura

A revisão da literatura é o meio pelo qual os pesquisadores identificam o conhecimento científico existente em uma determinada área de pesquisa [Mafra & Travassos 2006]. Com isso, eles podem planejar sua pesquisa, evitando a duplicação de esforços e a repetição de erros cometidos anteriormente.

No entanto, uma revisão da literatura conduzida sem um protocolo pré-estabelecido, pode ser dirigida por interesses pessoais de seus pesquisadores, levando a resultados pouco confiáveis, pouco abrangentes, não passíveis de repetição e, conseqüentemente, com pouco valor científico. Nesse sentido, a utilização de uma abordagem sistemática de revisão da literatura é fundamental para conduzir esse tipo de investigação.

De acordo com Biolchini et al. (2005), a revisão sistemática da literatura é uma metodologia de pesquisa desenvolvida com o objetivo de reunir e avaliar as evidências disponíveis em um determinado tema de pesquisa. Ela é um tipo de estudo secundário¹ e tem a capacidade de realizar revisões da literatura de forma rigorosa, informal e, até certo grau, repetível [Kitchenham & Charters 2005].

A metodologia de revisão sistemática da literatura possui como característica fundamental a definição de um **protocolo de revisão sistemática**, que especifica a questão de pesquisa, métodos e estratégias de busca utilizados, critérios para inclusão e exclusão de trabalhos, dentre outros. Além disso, suas estratégias de busca tentam detectar todas as evidências relevantes de um determinado tema de pesquisa e são documentadas para que os leitores possam avaliar o rigor, completude e o grau de repetibilidade do processo². Por fim, os critérios qualitativos usados para avaliar cada estudo primário relevante também são definidos e aplicados.

¹Estudos secundários são estudos que analisam todos os estudos primários relativos a uma determinada questão de pesquisa com o objetivo de integrar/sintetizar as evidências relacionadas a esta questão de pesquisa [Kitchenham & Charters 2005].

²É importante frisar que as pesquisas realizadas em bibliotecas digitais são quase sempre impossíveis de serem replicadas. No entanto, o processo em si pode ser replicado.

Essas características permitem que as revisões sistemáticas sejam utilizadas em diferentes situações como, por exemplo:

- Construção de um resumo (mapa) sobre as principais pesquisas e evidências relacionadas a um determinado tema de pesquisa;
- Identificação das eventuais lacunas (*gaps*) nas pesquisas atuais, com o objetivo de sugerir áreas que necessitam de uma investigação mais profunda;
- Avaliação do grau de profundidade em que as evidências empíricas suportam ou contradizem hipóteses teóricas;
- Auxílio na geração de novas hipóteses;

Nesse capítulo será apresentada uma revisão sistemática da literatura utilizada para construção de um mapa sobre sistemas de tradução automática para línguas de sinais, o tema de pesquisa foco deste trabalho. Esse processo foi conduzido por três pesquisadores: o autor deste trabalho e os pesquisadores Danilo Assis Nobre e Silva e Gutenberg Botelho Pessoa Neto³. A participação de mais de um pesquisador, permitiu, especialmente, que as etapas subjetivas do processo como, por exemplo, as fases de seleção, classificação e avaliação dos artigos fossem conduzidas de forma mais imparcial, tornando o processo mais rigoroso e os resultados mais confiáveis.

O protocolo de revisão sistemática desenvolvido para conduzir essa revisão é apresentado no Apêndice A. Na Seção 3.1, será apresentado o processo de busca, seleção e classificação dos trabalhos, resultados da aplicação desse protocolo. Na Seção 3.2, os principais trabalhos serão descritos e avaliados qualitativamente. Maiores detalhes sobre a metodologia de revisão sistemática podem ser encontrados em [Biolchini et al. 2005][Kitchenham & Charters 2005].

3.1 Busca, Seleção e Classificação dos artigos

A partir do protocolo definido no Apêndice A, inicialmente, foi aplicado um processo de busca por artigos científicos utilizando a "string de busca" e os portais de artigos definidos nesse protocolo. A consulta foi realizada no dia 13 de Setembro de 2011 e nela foram encontradas 493 referências distribuídas de acordo com a Tabela 3.1.

Um processo de seleção (filtragem) foi então aplicado nas referências encontradas. Esse processo foi conduzido pelos três revisores/avaliadores em duas iterações. Na primeira iteração, os revisores aplicaram os critérios de inclusão/exclusão em cada refe-

³Danilo e Gutenberg são discentes do Programa de Pós-Graduação em Informática da Universidade Federal da Paraíba (UFPB) e seus trabalhos de Mestrado envolvem as áreas de Acessibilidade e TV Digital.

rência, considerando-se apenas o título e o resumo do artigo. Na segunda iteração, os artigos que foram considerados indeterminados por, pelo menos, dois revisores, foram novamente revisados, considerando-se dessa vez a leitura completa do artigo. Após essas duas iterações, foram selecionados 56 artigos dentre as 493 referências encontradas. A lista completa com os 56 artigos selecionados está disponível no Apêndice B. A Tabela 3.1 apresenta os resultados do processo de busca e seleção das referências.

Tabela 3.1: Resultados do processo de busca e seleção de artigos

	IEEEExplore	ACM	Scopus	ScienceDirect	ISI	Total
Processo de busca	133	29	261	14	56	493
Seleção (1 ^a iteração)	26	16	23	4	3	72
Seleção (2 ^a iteração)	19	12	19	3	3	56

Uma das principais razões para o grande número de artigos eliminados na fase de seleção foi o elevado número de referências encontradas relacionados a tradução automática de línguas de sinais para línguas orais (isto é, a tradução no sentido inverso ao abordado nesse trabalho que envolve tradução de línguas orais para línguas de sinais). Além disso, muitos trabalhos foram referenciados por mais de um portal, o que resultou num grande número de documentos repetidos.

Após a seleção, os artigos selecionados passaram por um processo de classificação. Nesse processo, os três revisores/avaliadores leram os artigos e classificaram-nos com base nos critérios de classificação definidos no Apêndice A. Quando havia discordância entre os revisores na classificação de um artigo, os revisores se reuniam e discutiam suas posições até que um consenso fosse alcançado. O resultado geral do processo de classificação é apresentado na Tabela 3.2.

De acordo com a Tabela 3.2, considerando o método de tradução utilizado no trabalho, pode-se observar que a maioria dos artigos avaliados (40 artigos) utiliza Tradução Baseada em Regras, o que representa aproximadamente 71% dos artigos avaliados. Dentre os demais artigos, 5 deles são baseados no em Tradução Automática Estatística, 3 utilizam abordagens Híbridas (isto é, utilizam mais de um método de tradução) e 8 foram considerados "Sem classificação"(isto é, indicavam que faziam tradução, mas não explicavam como essa tradução era feita).

A preponderância de trabalhos baseados em Tradução Baseado em Regras pode ser explicada pela dificuldade em se encontrar corpus textuais em línguas de sinais, especialmente devido a ausência de um sistema de escrita formal e reconhecido para essas línguas.

Tabela 3.2: Classificação dos artigos de acordo com as categorias

Categoria	Subcategoria	Número de artigos	Percentual
Estratégia	Baseada em Regras	40	71,43%
	Estatística	5	8,93%
	Baseada em Exemplos	0	0,00 %
	Híbrida	3	5,36%
	Sem classificação	8	14,29%
Entrada	Voz	15	25,79%
	Texto	37	66,07%
	Sem classificação	4	7,14%
Saída	Animação	40	71,43%
	Vídeos gravados	2	3,57%
	Glosa	10	17,85%
	Representação escrita em LS	1	1,79%
	Sem classificação	3	5,36%
Domínio da Aplicação	Geral	33	58,93%
	Específico	20	35,71%
	Sem classificação	3	5,36%
Método de Avaliação	Automático	7	12,50%
	Com usuários	9	16,07%
	Híbrido	8	14,29%
	Sem avaliação	32	57,14%
Presença de Corpus	Sem corpus	44	78,57%
	Corpus Unilíngue (língua oral)	3	5,36%
	Corpus Unilíngue (língua de sinais)	0	0,00%
	Corpus Bilingue	9	16,07%
Tipo de artigo	Survey	0	0,00%
	Overview	0	0,00%
	Experimental	2	3,57%
	Sistema, descrição, modelo, etc	54	96,43%

Uma vez que os demais métodos (Tradução Automática Estatística e Tradução Baseada em Exemplos) são dependentes de informações estatísticas extraídas de um corpus textual bilingüe, os pesquisadores precisam desenvolver seu próprio corpus em língua de sinais, uma tarefa que, em geral, consome muito tempo. Isso pode ser verificado pelo número reduzido de trabalhos que utilizam algum corpus textuais em língua de sinais (apenas 9 artigos de acordo com a Tabela 3.2). Na Tabela 3.3 é apresentada uma análise mais detalhada dos métodos de tradução abordados nos artigos.

Tabela 3.3: Classificação dos artigos de acordo com os métodos de Tradução Automática

Estratégia	Número de artigos	Percentual
Baseada em Regras	40	71,43%
Tradução direta	14	25,00%
Transferência	20	35,71%
Interlíngua	4	7,14%
Híbridas ¹	2	3,57%
Estatística	5	8,93%
Baseado em palavras	2	3,57%
Baseado em frases	1	1,79%
Baseado em sintaxe	1	1,79%
Transdutores de estados finitos	0	0,00%
Híbridos ¹	1	1,79%
Baseada em Exemplos	0	0,00%
Híbrida ²	3	5,36%
Regras + Estatística	2	3,57%
Regras + Exemplos	0	0,00%
Estatística + Exemplos	0	0,00%
Regras + Estatística + Exemplos	1	1,79%
Sem classificação	8	14,29%

¹Que envolvem mais de um subtipo

²Que envolvem mais de um método de tradução (Regras, Estatístico e Exemplos)

De acordo com a Tabela 3.3, dentre os trabalhos que utilizam Tradução Baseada em Regras (40 artigos), a maioria deles utilizam a abordagem de Tradução Baseado em Transferência (20 artigos), isto é, uma estratégia que utiliza estruturas gramaticais nos níveis sintático e semântico para realizar a tradução entre as línguas. Dentre os demais artigos, 14 artigos utilizam abordagens de Tradução Direta, 4 utilizam abordagens baseadas em Interlíngua e 2 utilizam abordagens híbridas, isto é, que utilizam uma combinação de mais de um dos tipos de estratégias de Tradução baseada em Regras (Direta, Transferên-

cia e Interlíngua). Os trabalhos que envolvem apenas Tradução Direta, em geral, fazem apenas uma associação direta entre a palavra na língua oral e o sinal na língua de sinais, isto é, sem nenhuma adaptação para gramática da língua de sinais, e, em virtude disso, estão mais relacionados a animação/síntese de sinais (geração de vídeos em línguas de sinais a partir de agentes animados virtuais - avatares) do que com a tradução (adaptação gramatical) para língua de sinais propriamente dita.

Os trabalhos que utilizam Tradução Automática Estatística (5 artigos) utilizam modelos de tradução baseados em palavras (2 artigos) [Knight 1997], modelos de tradução baseados em frases (1 artigo) [Koehn et al. 2003] e modelos de tradução baseados em sintaxe (1 artigo) [Knight 1997]. O artigo restante, proposto por Gallo et al. (2009) e classificado como híbrido, envolve a tradução baseada em dois tipos de modelo de tradução: o modelo baseada em frases e um modelo baseado em transdutores de estados finitos [Kumar et al. 2004].

Os trabalhos classificados como de estratégia Híbridas (3 artigos) envolvem a combinação de mais de um tipo de estratégia de tradução (Regras, Estatístico e Exemplos). Os 3 artigos foram propostos pelos mesmos autores, [Lopez-Ludena et al. 2011][San-Segundo et al. 2007] e [San-segundo et al. 2008b] e envolvem tanto a combinação de estratégias de Tradução Baseada em Regras com Tradução Automática Estatística [San-Segundo et al. 2007][San-segundo et al. 2008b], como a combinação dos três tipos de estratégias (Regras, Estatística e Exemplos) na mesmo sistema de tradução [Lopez-Ludena et al. 2011].

Considerando o tipo de entrada e de saída utilizados na estratégia de tradução, pode-se observar que a maioria dos trabalhos explora a tradução a partir de textos na língua oral (37 trabalhos - 66,07%) e gera na saída animações representadas por avatares (40 trabalhos - 71,43%). No entanto, um número representativo de trabalhos utiliza voz como fonte de entrada das abordagem de tradução (15 trabalhos - 25,79%) ou geram saídas baseadas em glosa (10 trabalhos - 17,85%).

Com relação ao domínio da aplicação, 33 trabalhos (58,93%) projetaram sua estratégia de tradução pode ser aplicada em domínios gerais e abertos, enquanto que 20 trabalhos (35,71%) foram projetados para um domínio de tradução específico e restrito. Esses dados contrastam, no entanto, com o elevado número de trabalhos (32 trabalhos - 57,14%) que não descreveram ou aplicaram nenhum tipo de avaliação sobre a viabilidade, qualidade e desempenho da tradução (apresentados como "Sem classificação" na categoria Método de Avaliação). A Tabela 3.4 apresenta o tipo de avaliação utilizado para os 33 trabalhos de domínio geral.

Conforme pode ser observado na Tabela 3.4, embora muitos trabalhos tenham pro-

Tabela 3.4: Tipo de avaliação utilizada nos trabalhos de domínio geral (33 artigos)

Tipo de Avaliação	Número de Trabalhos	Percentual
Automático	1	3,03%
Com usuários	4	12,12%
Híbrido	3	9,09%
Sem avaliação	25	75,76%

posto abordagens de tradução de domínio geral, a maioria absoluta deles (25 trabalhos - 73,53%) não fizeram qualquer tipo de avaliação da solução, o que inviabiliza a obtenção de conclusões a respeito da qualidade, viabilidade e desempenho dessas soluções. Dentre os restantes, 5 trabalhos conduziram apenas avaliações com usuários, 1 utilizou apenas avaliação automática e 3 fizeram avaliações híbridas, isto é, avaliação automática e com usuários.

Além disso, outro dado importante que pode ser observado na Tabela 3.2 é que apenas 17 dos trabalhos considerados (30,36%) realizaram algum tipo de avaliação com usuários (somando os trabalhos que realizaram avaliações com usuários e os que realizaram avaliações híbridas). Esse tipo de avaliação com usuários é, em geral, bastante custosa e demorada, mas considerando que os surdos são o público alvo desse tipo de pesquisa, a sua participação é fundamental para avaliar a viabilidade, a qualidade e o grau de adequabilidade da solução por usuários surdos. De acordo com Su & Wu (2009), por exemplo, a avaliação objetiva com base em testes automáticos não é suficiente para avaliar a qualidade da tradução para línguas de sinais, uma que essas línguas são línguas gestuais e visuais.

Por fim, com relação ao tipo do artigo, a maioria absoluta dos artigos, 54 artigos (96,43%) propunham algum sistema, arquitetura, modelo, etc. para tradução automática para línguas de sinais. Os 2 trabalhos restantes (3,57%) reportavam estratégias usadas para avaliação de sistemas de tradução e foram classificados como "Experimentais". Na próxima seção será apresentada uma descrição e uma avaliação qualitativa dos principais trabalhos, destacando suas características, principais contribuições e resultados obtidos.

3.2 Avaliação Qualitativa das Principais Contribuições

Na Seção 3.1, um conjunto de artigos foi selecionado e classificado de acordo com base em um conjunto de critérios: estratégia de tradução utilizada, tipo de entrada e saída, domínio da aplicação, método de avaliação, presença de corpus e tipo de artigo. Dentre

os artigos selecionados, uma importante característica observada foi o elevado número de artigos (32 artigos - 57,14%) que não descreveu ou aplicou qualquer tipo de avaliação sobre a solução (ver Tabela 3.2). Conforme mencionado, essa falta de testes e experimentos sobre as soluções, inviabiliza a obtenção de conclusões a respeito da qualidade, viabilidade e desempenho dessas soluções e, conseqüentemente, dificulta a avaliação das contribuições do trabalho.

Além disso, também foi observado um número razoável de trabalhos (14 artigos - 25,00%) que utilizam apenas estratégias de tradução direta, isto é, que fazem apenas uma associação direta entre a palavra na língua oral e o sinal na língua de sinais, sem nenhuma adaptação para gramática da língua de sinais. Conforme mencionado, esses trabalhos estão mais relacionados com a sinalização (síntese/animação) das palavras na língua oral do que com a tradução automática para as línguas de sinais propriamente dita e, portanto, não são o foco principal dessa revisão sistemática.

Em virtude disso, os artigos que envolvem alguma estratégia de avaliação e que não utilizam apenas tradução direta foram selecionados (19 artigos) para uma avaliação qualitativa das suas principais contribuições e dos principais resultados obtidos. A lista desses artigos (em ordem alfabética) e um resumo das suas características são apresentados na Tabela 3.5. Nas próximas subseções, será apresentada uma descrição resumida e uma avaliação qualitativa de cada um desses trabalhos. Para organizar a apresentação desses trabalhos, eles foram organizados em subseções de acordo com a língua de sinais envolvida.

3.2.1 Tradução Automática para Língua Indiana de Sinais

Anuja et al. (2009) propuseram um sistema para traduzir automaticamente voz em Inglês para Língua Indiana de Sinais (*Indian Sign Language - ISL*) com enfoque em auxiliar pessoas surdas na interação em locais públicos como ferrovias e bancos (domínio específico). O sistema utiliza uma abordagem baseada em transferência para tradução das entradas de voz em inglês para animações em ISL. Inicialmente, um módulo Reconhecedor de Voz recebe o fluxo de entrada e o converte para uma seqüência de palavras (texto). Esse texto é então encaminhado para um módulo de Processamento de Linguagem que transforma as sentença do texto em uma estrutura gramatical (*phrase grammar structure*) no nível sintático e aplica um conjunto de regras de tradução, gerando uma seqüência de glosas. Por fim, um módulo de Animação-3D cria as animações para essa seqüência de glosas usando um avatar-3D.

A qualidade das animações geradas pela solução de Anuja et al. (2009) foi avaliada por

Tabela 3.5: Artigos que envolvem avaliação com usuários

Ref	Estratégia	Entrada	Saída	Domínio	Corpus	Língua Oral	Língua de Sinais
[Anuja et al. 2009]	Regras	Voz	Animação	Específico	Bilingue	Inglês	ISL
[Chiu et al. 2007]	Estatística	Texto	Vídeos gravados	Geral	Bilingue	Chinês	TSL
[Dangsaart et al. 2008]	Regras	Texto	S.C. ¹	Específico	Sem corpus	Tailandês	ThaiSL
[Dasgupta & Basu 2008]	Regras	Texto	Glosa	Geral	Sem corpus	Inglês	ISL
[D'Haro et al. 2008]	Estatística	Voz	Glosa	Específico	Unilíngue (Oral)	Espanhol	LSE
[Fotinea et al. 2008]	Regras	Texto	Animação	Geral	Sem corpus	Grego	GSL
[Huenerfauth et al. 2007]	S.C. ¹	S.C. ¹	Animação	Específico	Sem corpus	Inglês	ASL
[Huenerfauth 2008]	Regras	Texto	Animação	Geral	Sem corpus	Inglês	ASL
[Gallo et al. 2009]	Estatística	Voz	Animação	Específico	Bilingue	Espanhol	LSE
[Kim et al. 2004]	Regras	Texto	Animação	Geral	Sem corpus	Coreano	KSL, JSL e CSL
[Lopez-Ludena et al. 2011]	Híbrida	Voz	Animação	Específico	Bilingue	Espanhol	LSE
[Murakami et al. 2002]	S.C. ¹	Texto	Animação	S.C. ¹	Sem corpus	Japonês	JSL
[Sagawa et al. 1996]	S.C. ¹	Texto	Animação	Específico	Sem Corpus	Japonês	JSL
[San-Segundo et al. 2006]	Regras	Voz	Animação	Específico	Bilingue	Espanhol	LSE
[San-Segundo et al. 2007]	Híbrida	Voz	Glosa	Específico	Bilingue	Espanhol	LSE
[San-segundo et al. 2008a]	Regras	Voz	Animação	Específico	Sem Corpus	Espanhol	LSE
[San-segundo et al. 2008b]	Híbrida	Voz	Animação	Específico	Bilingue	Espanhol	LSE
[Su & Wu 2009]	Estatística	Texto	Glosa	Geral	Bilingue	Chinês	TSL
[Wu et al. 2007]	Estatística	Texto	Glosa	Geral	Bilingue	Chinês	TSL

¹ S.C.: Sem classificação

usuários surdos a partir de um conjunto de testes simples com 250 sentenças no domínio do problema (bancos e ferrovias). Nesse teste, usuários avaliaram essas frases e definiam se a tradução estava plenamente correta (60%), se possuía erros semânticos (21%) ou se era incompreensível (19%). Contudo, o número de usuários utilizados no teste, bem como os critérios usados para classificar cada sentença nessas três classes e as informações estatísticas do teste não foram descritas pelos autores. Além disso, de acordo com os autores, o desempenho da solução com relação ao tempo de tradução não é aceitável, isto é, o tempo para geração das traduções ainda é muito alto (os valores, no entanto, não foram informados).

A solução proposta por Dasgupta & Basu (2008) envolve a tradução automática de textos em Inglês para glosas em ISL. A estratégia de tradução utilizada é a tradução baseada em transferência (regras) que utiliza estruturas gramaticais denominadas *lexical functional grammars f-structure* (LFG *f-structure*) para representar as informações morfológicas e sintáticas dos elementos da sentença. O sistema inicia executando um pré-processamento no texto de entrada com o objetivo de converter as sentenças em uma estrutura do tipo *f-structure*. As sentenças em *f-structure* são então convertidas para uma seqüência de glosas em ISL a partir de duas operações principais: (1) seleção léxica e a (2) reordenação das palavras. A seleção léxica é executada a partir da busca em um dicionário bilingüe que relaciona cada palavra em inglês com uma glosa em ISL. A reordenação das palavras faz os ajustes na ordem de apresentação das palavras, gerando uma seqüência de glosas em ISL. A arquitetura do sistema também prevê um módulo para síntese de sinais a partir de uma seqüência de glosas, mas, segundo os autores, esse módulo ainda não foi desenvolvido. A avaliação da solução foi realizada com a participação de usuários surdos que avaliaram a qualidade e compreensão das saídas geradas pela solução para 208 sentenças. Contudo, de forma similar ao trabalho proposto por Anuja et al. (2009), o número de usuários utilizados no teste, os critérios usados para avaliar as sentenças e as informações estatísticas do teste não foram descritas pelos autores.

3.2.2 Tradução Automática para Língua Taiwanesa de Sinais

Chiu et al. (2007) propuseram uma estratégia para tradução automática estatística de textos em Chinês para Língua Taiwanesa de Sinais (*Taiwanese Sign Language* - TSL) com a saída sendo gerada a partir da síntese de vídeos gravados por intérpretes. A estratégia de tradução utiliza um corpus bilingüe em Chinês e TSL que contém, além dos textos paralelos, informações lingüísticas e informações de sinalização relacionadas aos textos. O alinhamento entre os trechos do corpus é realizado em dois níveis hierárquicos, um no

nível morfológico e outro no nível sintático, de acordo com as informações morfológicas e sintáticas presentes no corpus. Isso permite que a complexidade do modelo de tradução seja reduzido, tornando a tradução viável para corpus bilingües pequenos. Além disso, o corpus também armazena algumas informações de sinalização, como, por exemplo, a localização inicial e final das mãos em cada sinal. Uma vez que a saída do sistema (vídeo em TSL) é gerada a partir da síntese (combinação) de vídeos gravados por intérpretes, a estratégia de suavização utiliza as informações relacionadas a localização das mãos para suavizar a transição entre sinais (ou vídeos) consecutivos.

Na versão do corpus desenvolvida pelos autores, 1983 sentenças em Chinês foram alinhadas com 1983 sentenças em TSL contendo as informações morfo-sintáticas. Além disso, as localizações das mãos também foram anotadas no corpus e foram identificados 891 gestos (sinais) que envolvem as 2 mãos, 417 que envolvem apenas a mão direita e 8 que envolvem a mão esquerda.

A avaliação dessa solução foi conduzida usando testes objetivos e avaliações com usuários. Nos testes objetivos, a qualidade da tradução foi comparada com o modelo **IBM Model 2** usando métricas objetivas. Esses testes indicaram que a solução proposta por Chiu et al. (2007) obteve uma melhor qualidade de tradução. Os testes com usuários foram conduzidos com 5 usuários surdos que avaliaram a naturalidade dos vídeos gerados pela solução, com base em 4 diferentes estratégias de suavização. O tempo médio para gerar as traduções, no entanto, não foi avaliado.

Su & Wu (2009) e Wu et al. (2007) também propuseram estratégias de tradução automática estatística de textos em Chinês para glosas em TSL. Essas estratégias derivam gramáticas livres de contextos probabilísticas (*Probabilistic Context-Free Grammars - PCFGs*) [Wu et al. 2007] e gramáticas livres de contexto síncronas (*Synchronous Context Free Grammar - SCFG*) [Su & Wu 2009] a partir de um corpus paralelos bilingüe com informações sintáticas e semânticas, com o objetivo de extrair memórias de tradução sintáticas⁴. De forma similar a proposta de Chiu et al. (2007), a idéia desses trabalhos é construir tradutores eficientes utilizando um conjunto menor de dados (corpus menores). Um dos problemas dessas abordagens, no entanto, é que o corpus precisa estar anotado com as informações sintáticas nos dois corpus.

As duas soluções (Su & Wu (2009) e Wu et al. (2007)) foram avaliadas usando testes objetivos e avaliações com usuários. Nos testes objetivos das duas soluções, as métricas objetivas de WER (*Word Error Rate* ou Taxa de Erro de Palavra) e BLEU (*Bilingual Eva-*

⁴As memórias de tradução representam bases de dados que contém fragmentos traduzidos em duas línguas. Em Wu et al. (2007) e [Su & Wu 2009], a memória de tradução representa pares de regras gramaticais envolvendo as línguas Chinesa e TSL.

luation Understudy [Papineni et al. 2001], algumas das métricas objetivas mais utilizada para avaliar sistemas de tradução, foram aplicadas para avaliar a qualidade da tradução em um conjunto de sentenças. Os resultados mostraram que estas soluções conseguiram obter melhores qualidade de tradução que modelos tradicionais de tradução automática estatística, **IBM Model 3** [Wu et al. 2007] e **IBM Model 2** [Su & Wu 2009], de forma similar a Chiu et al. (2007). Nos testes com usuários, as soluções foram avaliadas usando testes MOS (*Mean Opinion Score*). Nesses testes usuários avaliam os conteúdos gerados pela solução usando escalas que variam entre ruim, razoável e bom. Os resultados apresentados pelos autores mostram que nos dois trabalhos, a maioria dos usuários considerou a qualidade das sentenças traduzidas boas ou razoáveis e obteve bom índice de compreensão dos conteúdos. O tempo de tradução dessas soluções, no entanto, não foram avaliados.

3.2.3 Tradução Automática para Língua Tailandesa de Sinais

Dangsaart et al. (2008) propuseram um sistema de tradução de textos em Tailandês para Língua Tailandesa de Sinais (*Thai Sign Language - ThaiSL*) com enfoque em educação (domínio específico). O sistema utiliza uma abordagem de tradução baseada em transferência e é composto por vários módulos. Inicialmente, o sistema passa por uma etapa de pré-processamento de texto, onde a sentença é quebrada em um conjunto de palavras (*tokens*). Em seguida, o sistema aplica uma substituição léxica das palavras pelas correspondentes glosas com base em um dicionário de glosas em ThaiSL, seguida por uma etapa de ordenação dos elementos da sentença para adaptação a gramática da ThaiSL. Por fim, as glosas reordenadas são convertidas para um conjunto de imagens estáticas. Uma das limitações desse trabalho, portanto, é que, em vez de animações ou vídeos, imagens estáticas são utilizadas para representar os sinais em ThaiSL.

O trabalho foi avaliado usando testes objetivos e avaliações com usuários. Nos testes objetivos, o desempenho de tradução foi avaliado com base em métricas objetivas (Acurácia, Precisão e *Recall*) e comparado com o modelo **IBM Model 2** usando métricas objetivas. Esses testes indicaram que a solução proposta por Dangsaart et al. (2008) obteve uma melhor qualidade de tradução. Os testes com usuários foram conduzidos através de questionários com o objetivo de avaliar o grau de satisfação dos usuários com os conteúdos gerados, considerando uma escala de 1 a 5, onde 5 representa "excelente" e 1 representa "péssimo". Em ambos os testes (objetivos e subjetivos) os resultados obtidos foram satisfatórios. Nesses testes, o tempo médio para gerar os conteúdos traduzidos para ThaiSL também não foram mensurados.

3.2.4 Tradução Automática para Língua Espanhola de Sinais

Lopez-Ludena et al. (2011), Gallo et al. (2009), San-Segundo et al. (2006), San-segundo et al. (2008a) e San-segundo et al. (2008b) propuseram uma arquitetura para traduzir voz em língua espanhola de sinais (LSE) com enfoque em ajudar pessoas surdas no atendimento em serviços públicos como, por exemplo, na renovação da carteira nacional de habilitação (domínio específico). A idéia do sistema é facilitar o diálogo entre surdos e prestadores de serviços públicos nesse tipo de serviço.

A arquitetura da solução é composta por três módulos: um Reconhecedor de Voz, um Tradutor de Linguagem Natural e um módulo de Animação. O Reconhecedor de Voz é utilizado para converter a fala em uma seqüência de palavras (texto). O módulo Tradutor de Linguagem Natural converte a seqüência de palavras em uma seqüência de sinais em LSE e, por fim, o módulo de Animação apresenta a seqüência de sinais usando um avatar-3D. A diferença básica entre esses trabalhos ([Lopez-Ludena et al. 2011][Gallo et al. 2009], [San-Segundo et al. 2006], [San-segundo et al. 2008a] e [San-segundo et al. 2008b]) é a estratégia de tradução utilizada pelo módulo Tradutor de Linguagem Natural.

Em [San-Segundo et al. 2006] e [San-segundo et al. 2008a], o módulo de Tradução foi proposto com base em uma estratégia de tradução baseada em transferência (regras), onde as regras de tradução eram definidas por especialistas. Mais especificamente, nesse processo de tradução, inicialmente as palavras são mapeadas para um ou mais rótulos sintáticos e pragmáticos [San-Segundo et al. 2006] ou semânticos [San-segundo et al. 2008a] e, em seguida, um conjunto de regras de tradução é aplicado para converter as palavras rotuladas em um conjunto de gestos.

Em [San-segundo et al. 2008b], o módulo de Tradução foi proposto com base em duas estratégias: uma estratégia baseada em regras (similar a estratégia proposta em [San-segundo et al. 2008a]), e uma estratégia baseada em tradução automática estatística. A estratégia de tradução automática estatística foi projetada com base no modelos de tradução baseados em frases [Koehn et al. 2003] e utiliza corpus paralelo bilingüe para treinar os modelos de linguagem e de tradução. O corpus paralelo bilingüe foi desenvolvido pelos próprios autores do trabalho e consiste de um conjunto de 416 sentenças dentro do domínio do problema (agentes públicos durante o processo de renovação de carteira de motorista). Nesse trabalho, a estratégia de tradução baseada em regras obteve melhores resultados do que a estratégia de tradução automática estatística. Segundo os autores isso pode ser explicado porque o domínio de tradução era restrito o que facilitava o desenvolvimento de um conjunto completo de regras com um pequeno esforço. Além disso, o corpus era muito pequeno, o que inviabilizava um treinamento apropriado dos modelos

estatísticos.

Em [Gallo et al. 2009], o modelo de tradução foi proposto com base em duas estratégias de tradução automática estatística. A primeira estratégia adotando modelos de tradução baseados em frases, de modo similar a solução proposta em San-segundo et al. (2008b), e a segunda adotando modelos de tradução baseados em transdutores de estados finitos [Kumar et al. 2004].

Em [Lopez-Ludena et al. 2011], uma estratégia híbrida que combina tradução baseada em regras, tradução automática estatística e tradução baseada em exemplos foi utilizada no módulo de tradução. Nessa estratégia, as estratégias são estruturadas de forma hierárquica. Inicialmente, uma tradução baseada em exemplos é aplicada. Se a medida de distância entre a sentença de entrada e o exemplo mais similar for menor que um certo limiar, a saída gerada pelo sistema é o exemplo traduzido. Caso contrário, um tradutor baseado em regras é aplicado (similar ao proposto por [San-segundo et al. 2008a]). Se a relação entre o número de glosas geradas pelo tradutor baseado em regras e o número de palavras da sentença de entrada for maior que um determinado limiar, a tradução gerada pelo tradutor baseado em regras é a saída do sistema. Caso contrário, um processo de tradução automática estatística (similar ao proposto por Gallo et al. (2009)) é aplicado.

Embora estes trabalhos (Lopez-Ludena et al. (2011), Gallo et al. (2009), San-Segundo et al. (2006), San-segundo et al. (2008a) e San-segundo et al. (2008b)) tenham investigado diferentes abordagens de tradução e produzido importantes contribuições pro campo da tradução automática para línguas de sinais, uma das limitações desses trabalhos é que além de serem restritos a um domínio específico, o menor tempo médio obtido para traduzir voz para LSE nessas soluções foi de, aproximadamente, 8 segundos por sentença, o que inviabiliza sua aplicação em cenários que exigem tradução em tempo real, como, por exemplo, na TV. A qualidade de tradução dessas soluções foi avaliada através de testes automáticos com medidas objetivas como, por exemplo, WER e BLEU [Papineni et al. 2001].

San-Segundo et al. (2007) propôs um trabalho experimental cujo objetivo era comparar a qualidade de tradução de diferentes abordagens de tradução automática de voz em espanhol para LSE propostas por San-segundo et al. (2008a), San-segundo et al. (2008b) e Gallo et al. (2009). Mais especificamente, três alternativas foram avaliadas: uma abordagem de tradução baseada em regras [San-segundo et al. 2008a], [San-segundo et al. 2008b], uma abordagem de tradução automática estatística baseada em frases [Gallo et al. 2009] e uma abordagem de tradução automática estatística baseada em transdutores de estados finitos [Gallo et al. 2009]. As estratégias foram comparadas com relação a quatro medidas objetivas: SER (*Sign Error Rate*), PER (*Position Independent Error Rate*),

BLEU (*BiLingual Evaluation Understudy*) e NIST.

Nessa avaliação, a estratégia de tradução baseada em regras também obteve os melhores resultados. No entanto, segundo os autores, o desenvolvimento das regras torna difícil a extensão para domínios gerais. Dentre as estratégias de tradução automática estatística, a estratégia baseada em transdutores de estados finitos obteve melhores resultados um pouco melhores, mesmo com um modelo que exige menos esforço pra ser desenvolvido.

Por fim, D'Haro et al. (2008) propôs uma técnica para criação de um novo modelo de linguagem para os tradutores estatísticos para LSE propostos por [Gallo et al. 2009]. A proposta do novo modelo é adaptar os tradicionais "modelos na linguagem de destino" utilizados na tradução automática estatística para assegurar que as sentenças traduzidas são bem formadas e fluentes na língua alvo. Para isso, o modelo interpola a informação obtida do modelo tradicional com contadores de frequência obtidos a partir da Web usando APIs do Google ⁵. Um conjunto de testes foi conduzido para medir a perplexidade⁶. Os resultados indicaram que o modelo adaptado proposto por D'Haro et al. (2008) obteve melhoria relativa de 18,9 % e 15,5 % na medida de perplexidade com relação ao modelo tradicional.

3.2.5 Tradução Automática para Língua Grega de Sinais

Fotinea et al. (2008) propuseram um sistema para tradução automática de textos em Grego para Língua Grega de Sinais (*Greek Sign Language - GSL*). O trabalho utiliza uma estratégia de tradução baseada em transferência (regras) para gerar as sentenças em GSL, mas o seu foco principal é na estratégia de animação que procura explorar a estrutura paralela das línguas de sinais, isto é, a capacidade apresentar um movimento das mãos com uma expressão facial. Para isso, um avatar-3D foi projetado com a opção de explorar a estrutura paralela das línguas de sinais.

Na tradução automática, inicialmente, um *parser* é aplicado sobre as sentenças de entrada, rotulando seus elementos com informações lingüísticas. Em seguida, uma estratégia de transferência é aplicada, convertendo as estruturas em Grego geradas pelo *parser* para uma estrutura similar em GSL. Essas operações de transferência envolvem, em geral, a adição, remoção ou a substituição dos elementos sintáticos. A estrutura em GSL é então encaminhada para um módulo de síntese que vai utilizar o avatar-3D para apresentar os sinais com base em um Dicionário em GSL (*GSL Lexicon*). Esse Dicionário que contém *scripts* que serão renderizados para gerar os sinais em GSL.

⁵<http://code.google.com/apis/ajaxsearch/>

⁶A perplexidade de uma distribuição de probabilidade está relacionada com a entropia da distribuição e é utilizado em Processamento de Linguagem Natural (PLN) para avaliar modelos de linguagem.

A avaliação da solução foi realizada por usuários em um domínio educacional. Para isso, o avatar-3D foi incorporado no protótipo de uma plataforma educacional de tutoria, e especialistas em GSL e educadores surdos avaliaram qualitativamente a usabilidade da solução, aceitabilidade do avatar-3D, dentre outros. O número de usuários utilizados no teste, os critérios usados para classificar para avaliar os aspectos e as informações estatísticas do teste, no entanto, não foram descritas pelos autores. O único resultado comentado pelo autor foi que o avatar obteve um desempenho satisfatório, mas foi considerado pouco natural. A qualidade o desempenho da tradução, por outro lado, não foram avaliados pelos autores.

3.2.6 Tradução Automática para Língua Japonesa de Sinais

Kim et al. (2004) propuseram um sistema para comunicação em tempo real entre surdos utilizando Língua Japonesa (*Japanese Sign Language - JSL*), Coreana (*Korean Sign Language - KSL*) e Chinesa de Sinais (*Chinese Sign Language - CSL*). A idéia do trabalho é desenvolver uma espécie de *chat* para comunicação entre surdos através de avatares. Uma vez que não existe um sistema de escrita reconhecido para línguas de sinais, o sistema utiliza sentenças em Koreano, Chinês ou Japonês como entrada e traduz essas sentenças para KSL, CSL ou JSL. De acordo com os autores, devido a similaridade entre as línguas KSL e JSL, a tradução entre as sentenças nessas línguas pode ser realizada através de tradução direta. Nesse caso, inicialmente, um cliente digita uma sentença de entrada e o servidor converte essa sentença em um conjunto de parâmetros de um avatar (codificação) de acordo com Dicionários em JSL ou KSL. Em seguida, esses parâmetros são encaminhados para um grupo de clientes que os recebe e gera uma animação usando avatares.

Segundo os autores, no entanto, a CSL possui uma estrutura lingüística diferente da JSL e KSL. Nesse caso, uma reordenação da sentença é aplicada pelo servidor usando um conjunto simples de regras de tradução no nível sintático (tradução baseada em transferência) antes de ser convertida no conjunto de parâmetros (codificação) que será transmitido através do canal de comunicação.

O trabalho, no entanto, apresenta algumas limitações com relação a avaliação da solução. Os únicos fatores avaliados foram o tempo de tradução e o número de bytes transmitidos pelo canal de comunicação e eles foram realizados apenas para uma única sentença, a sentença "Eu vou para a escola". Para essa sentença, a sentença foi traduzida em 188 milissegundos e foram transmitidos 826 bytes.

Com o objetivo de tornar as animações geradas pelos sistemas de tradução automática

para línguas de sinais mais naturais, Murakami et al. (2002) propuseram uma estratégia para inclusão de modificadores em um sistema de tradução de japonês para JSL. Os modificadores são utilizados nas línguas de sinais para realçar determinadas características. Por exemplo, na sentença "O livro está muito barato", os surdos naturalmente não representam a palavra "muito". Em vez disso, eles expressam o sinal "barato" com gestos exagerados para transmitir o significado de "muito". A solução, portanto, procurar incorporar esses modificadores nas sentenças com o objetivo de tornar a animação mais natural.

Mais especificamente, o método proposto utiliza regras gramaticais com possíveis variações em três características na animação do sinal: Extensão, Velocidade e Frequência. Por exemplo, em uma expressão do tipo "muito longe", a palavra "muito" duplica a extensão do sinal "longe". Nesse caso, quando a expressão "muito longe" aparece em uma sentença, o sistema produz uma animação dessa expressão usando o sinal "longe" com o parâmetro "Extensão" com valor 2. Isso também pode ser aplicado para os modificadores Velocidade e Frequência. Por exemplo, na expressão "muito rápido", a palavra "muito" duplica a velocidade de apresentação do sinal "rápido". Nesse caso, quando a expressão "muito rápido" aparece em uma sentença, o sistema produz uma animação dessa expressão usando o sinal "rápido" com o parâmetro "Velocidade" com valor 2. Essa estratégia permite que os dicionários sejam menores, uma vez que não é necessário registrar uma animação para cada expressão enfatizada.

A solução foi avaliada através de testes com 23 usuários que incluíam 5 usuários portadores de deficiência auditiva e 18 especialistas em JSL. Os usuários foram convidados a avaliar duas animações através de questionários, onde a primeira animação não utilizava modificadores e a segunda animação utilizava modificadores gerados pela solução de Murakami et al. (2002). De acordo com os autores, a animação com modificadores foi considerada mais natural pelos avaliadores. Os valores resultantes dessa avaliação e os dados estatísticos, no entanto, não foram apresentados pelos autores, o que dificulta a avaliação do grau de confiança dessa avaliação.

Sagawa et al. (1996) propuseram um sistema que explora a tradução automática entre Japonês e JSL nos dois sentidos: a tradução de JSL para Japonês e a tradução de Japonês para JSL. O subsistema de tradução de Japonês para JSL (tema desse trabalho), funciona de forma similar a maioria das soluções apresentadas anteriormente. Inicialmente, uma sentença de entrada em Japonês é convertida para uma sequência de glosas dos sinais em JSL (a estratégia de tradução, no entanto, não é descrita no artigo e, por isso, sua estratégia de tradução foi categorizada como "Sem classificação"). Após a geração das glosas, um componente de síntese obtém essas glosas e procura os padrões de animação de cada sinal

em um dicionário de sinais. Esses padrões são então animados com base no modelo de um avatar-3D, gerando as animações em JSL.

Um protótipo do sistema foi desenvolvido para auxiliar surdos na operação de caixas automáticas ATM (*Automatic Teller Machine*). Esse protótipo possui apenas 11 sentenças em JSL e foi avaliado por 4 usuários surdos e 2 intérpretes de JSL. De acordo com os autores, todos os usuários conseguiram compreender as sentenças, embora alguns deles tenham tido dificuldades em algumas sentenças. Ainda de acordo com os autores, a dificuldade pode ser explicada pela dificuldade em interpretar alguns movimentos manuais e pelas limitações nas expressões faciais.

3.2.7 Tradução Automática para Língua Americana de Sinais

Huenerfauth (2008) propôs uma abordagem híbrida de tradução automática baseada em regras, onde as sentenças em Inglês são traduzidas para Língua Americana de Sinais (*American Sign Language - ASL*) combinando uma estratégia de tradução baseada em interlíngua, uma estratégia de tradução baseada em transferência e uma estratégia de tradução direta, que são aplicadas de acordo com a sentença de entrada. Outra característica importante da solução de Huenerfauth (2008), é a capacidade de modelar classificadores no sistema de tradução.

Os classificadores são fenômenos lingüísticos utilizados pelos intérpretes para tornar as sinalizações mais naturais e mais fáceis de compreender. Mais especificamente, eles exploram o caráter gestual das línguas de sinais para representar informações, como, por exemplo, marcações topológicas de uma entidade imaginária no espaço de sinalização ou o movimento e reposicionamento de um objeto imaginário criado com o objetivo de indicar uma localização, movimento ou forma desse objeto.

Por exemplo, na sentença "O carro foi estacionado entre o gato e a casa", é possível utilizar três classificadores: (1) a mão não dominante representando um "C para baixo" para indicar uma posição imaginária no espaço para uma casa; (2) a mão dominante representando um "V inclinado para baixo", para indicar uma posição imaginária no espaço para o gato e (3) a mão dominante representando um "3 lateralmente", traçando um caminho no espaço correspondendo ao carro dirigindo e parando entre as posições espaciais da "casa" e do "gato".

Antes de utilizar cada um desses classificadores, no entanto, o sinalizador precisa representar um sinal para indicar que objeto está sendo descrito com o classificador. No exemplo acima, por exemplo, o sinalizador precisaria representar os sinais "CASA", "GATO" e "CARRO", respectivamente, antes de representar cada um dos classificadores

listados acima.

Para representar classificadores nos sistemas de tradução, portanto, além de uma análise sintática e semântica, é necessário também explorar conceitos como analogia espacial, visualização de cena, dentre outros. Em outras palavras, uma abordagem baseada em transferência (que explora apenas informações sintáticas e semânticas), não é suficiente para representar esses classificadores. Dessa forma, é necessário explorar a tradução em um nível semântico mais profundo, isto é, é necessário explorar o uso de abordagens de tradução baseados em interlíngua.

Um dos problemas no uso de soluções baseadas em interlíngua, no entanto, é a dificuldade de projetar um sistema baseada em interlíngua para um domínio geral [Huenerfauth 2008]. Em geral, esse tipo de abordagem só é viável para domínios limitados. Além disso, existem sentenças em Inglês em que uma análise semântica profunda não é necessária. Nesse caso, uma abordagem baseada em transferência poderia ser utilizada para essas sentenças. Explorando essas idéias, a solução proposta por Huenerfauth (2008) propôs uma arquitetura híbrida que envolve tanto tradução baseada em interlíngua, como tradução baseada em transferência e tradução direta.

Mais especificamente, nessa arquitetura, a estratégia de tradução aplicada depende da sentença de entrada. Quando a sentença pertence ao domínio restrito da tradução baseada em interlíngua, um processo de tradução baseada em interlíngua (com o uso de classificadores) é aplicada na sentença. Quando a sentença não pertence a esse domínio, mas possui informações lingüísticas que são contempladas pelo conjunto de regras de tradução definidas na tradução baseada em transferência, então um processo de tradução baseada em transferência é aplicado na sentença. Por fim, quando a sentença não se encaixa em nenhuma das duas situações anteriores, uma tradução direta é aplicada na sentença.

Huenerfauth et al. (2007) conduziu um experimento com usuários para confrontar a qualidade dos conteúdos gerados pela solução baseada em classificadores [Huenerfauth 2008] com a qualidade de conteúdos gerados por uma solução baseada em tradução direta. Os resultados mostraram que os usuários consideraram os conteúdos gerados pela solução de Huenerfauth (2008) significativamente mais naturais, gramaticalmente corretos e compreensíveis do que a solução baseada em tradução direta. Além disso, em testes realizados para avaliar o nível de compreensão dos usuários foi possível verificar também que os usuários conseguiram compreender melhor os conteúdos quando o conteúdo explorava o uso de classificadores [Huenerfauth 2008]).

3.2.8 Avaliação dos Principais Resultados

Analisando os trabalhos discutidos nessa seção, pode-se observar que as estratégias de tradução automática propostas e utilizadas nesses trabalhos são bastante sofisticadas e diversificadas. No entanto, também é possível observar a ausência de soluções que envolvam tradução automática em tempo real e para domínios gerais, como, por exemplo, na TV. Outro aspecto importante que pode ser observado, é que nenhum dos trabalhos avaliados explora algum tipo de estratégia de computação humana para auxiliar no desenvolvimento das construções lingüísticas das soluções, bem como na evolução da qualidade dos conteúdos gerados ao longo do tempo, o que dificulta a inserção de novas construções ou novos sinais ou a edição das construções existentes ao longo do tempo. Além disso, nenhum dos trabalhos avaliados aborda a tradução para Língua Brasileira de Sinais (LIBRAS), a língua de sinais oficial do Brasil (foco deste trabalho).

Por fim, os aspectos avaliados e as metodologias de avaliação utilizadas pelos trabalhos não são uniformes. Alguns trabalhos avaliam apenas a qualidade de tradução utilizando métricas objetivas, outros avaliam utilizando testes com usuários, mas sem descrever os valores obtidos ou a metodologia utilizada, além de muitos negligenciarem importantes aspectos desse tipo de solução, como o nível de compreensão dos conteúdos pelos usuários, o tempo ou atraso de tradução, a naturalidade da apresentação, dentre outros. A Tabela 3.6, apresenta um resumo dos aspectos avaliados em cada um dos trabalhos, com um 'X' indicando que o aspecto foi avaliado no trabalho. Para minimizar esse problema, no Capítulo 6 foi proposta uma metodologia para avaliação desse tipo de solução que contempla todos os aspectos apresentados na Tabela 3.6.

3.3 Considerações

Nesse capítulo foi apresentado um processo de revisão desenvolvido sobre sistemas de tradução automática para língua de sinais. Nesse processo foi definido, inicialmente um protocolo de revisão para direcionar o processo de busca, seleção e classificação dos trabalhos. Em seguida, o protocolo foi aplicado e um conjunto de trabalhos foi selecionado e classificado. Os principais trabalhos passaram então por um processo de descrição e avaliação qualitativa das principais contribuições e resultados obtidos.

Nessa avaliação qualitativa, pode-se perceber que embora as estratégias de tradução automática utilizadas nesses trabalhos sejam bastante sofisticadas e diversificadas, a maioria dos trabalhos carece de uma metodologia apropriada de avaliação da solução. Além disso, pode-se observar a ausência de trabalhos que explorem a tradução em tempo real

Tabela 3.6: Artigos que envolvem avaliação com usuários

Ref	Aspectos Avaliados				
	Compreensão	Qualidade Tradução (Objetivo)	Qualidade Tradução (Subjetivo)	Naturalidade Animação	Tempo Tradução (Tempo real?)
[Anuja et al. 2009]	-	-	X	-	X (Não)
[Chiu et al. 2007]	-	X	-	X	-
[Dangsaart et al. 2008]	-	X	X	-	-
[Dasgupta & Basu 2008]	X	-	X	-	-
[D'Haro et al. 2008]	-	X	-	-	-
[Fotinea et al. 2008]	-	-	-	X	-
[Huenerfauth et al. 2007]	X	X	-	X	-
[Huenerfauth 2008]	X	X	-	X	-
[Gallo et al. 2009]	-	X	-	-	X (Não)
[Kim et al. 2004]	-	-	-	-	X (Não)
[Lopez-Ludena et al. 2011]	-	X	-	-	X (Não)
[Murakami et al. 2002]	-	-	-	X	-
[Sagawa et al. 1996]	X	-	-	-	-
[San-Segundo et al. 2006]	-	X	-	-	X (Não)
[San-Segundo et al. 2007]	-	X	-	-	X (Não)
[San-segundo et al. 2008a]	-	X	-	-	X (Não)
[San-segundo et al. 2008b]	-	X	-	-	X (Não)
[Su & Wu 2009]	X	X	X	-	-
[Wu et al. 2007]	X	X	X	-	-

e para cenários que envolvem conteúdos de domínio geral como, por exemplo, a TV, e que explorem algum tipo de estratégias de computação que auxilie no desenvolvimento das construções lingüísticas das soluções. Por fim, também pôde-se observar que nenhum desses trabalhos aborda a Língua Brasileira de Sinais.

Essas limitações, somadas aos problemas de acesso dos surdos às TIC, motivaram o desenvolvimento a solução proposta nesse trabalho que será apresentada em detalhes no próximo capítulo.

Capítulo 4

Solução Proposta

Nesse capítulo, a arquitetura e os componentes da solução proposta serão apresentados. Conforme mencionado no Capítulo 1, a solução tem como objetivo reduzir os problemas de comunicação dos surdos em TIC, como, por exemplo, TV, Web e Cinema Digital, especialmente quando intérpretes humanos não estão disponíveis. Para isso, a solução é composta por um conjunto de componentes responsáveis por gerar automaticamente (isto é, sem intervenção humana direta) trilhas de LIBRAS em conteúdos multimídia a partir das legendas desses conteúdos.

Uma característica importante e inédita dessa solução é que ela inclui uma estratégia eficiente de tradução automática de textos em língua portuguesa para glosas em LIBRAS e utiliza dicionários de LIBRAS para minimizar os recursos computacionais necessários para gerar os conteúdos de LIBRAS em tempo real. A estratégia de tradução de texto para glosa foi projetada para traduzir conteúdos de forma eficiente e para domínios gerais e combina métodos de compressão estatística utilizados para classificar os *tokens* (palavras) de entrada, estratégias de simplificação textual para reduzir a complexidade do texto de entrada e um conjunto de regras morfológicas e sintáticas definido por especialistas.

Além disso, dicionários de LIBRAS são utilizados para evitar a renderização dos sinais em tempo real, uma vez que essa tarefa consome muito tempo. Esses dicionários armazenam vídeos dos sinais de LIBRAS pré-renderizados e cada sinal possui um código (por exemplo, sua representação textual em glosa) associado com esse vídeo. Dessa forma é possível gerar um vídeo de LIBRAS a partir da combinação de sinais no dicionário de LIBRAS.

Outro importante e inédito aspecto da solução é a utilização de estratégias de colaboração e computação humana para desenvolver as construções lingüísticas da solução de forma eficiente e semi-automática. A idéia dessa abordagem é que especialistas em LIBRAS colaborem na geração dessas construções lingüísticas e também melhorem a qualidade dos conteúdos gerados através da melhoria das regras de tradução, da inclusão

de novos sinais, etc. Para isso, uma ferramenta de computação humana, denominada WikiLIBRAS, foi desenvolvida, juntamente com linguagens formais para descrever regras de tradução (Linguagem de Descrição de Regras de Tradução) e sinais (Linguagens de Descrição de Sinais), e o modelo de um agente animado virtual 3D (avatar-3D).

Uma visão esquemática da solução proposta é apresentada na Figura 4.1. De acordo com a Figura 4.1, a solução funciona da seguinte forma. Inicialmente, os conteúdos multimídia submetidos a solução passam por um componente de Filtragem, responsável por extrair as trilhas de legendas desse conteúdos. Opcionalmente, um arquivo (ou fluxo) de legenda pode ser carregado diretamente na solução. Em seguida, um componente de Extração de Legendas converte esse fluxo (ou arquivo) de legenda em uma seqüência de palavras em língua portuguesa. Utilizando o componente de Tradução Automática, essa seqüência de palavras é então automaticamente traduzida para uma seqüência de glosas em LIBRAS. A seqüência de glosas é então enviada para um componente de Animação que associa cada glosa com uma representação visual de um sinal (vídeo) no Dicionário de LIBRAS. Dessa forma, a seqüência de glosas é mapeada para uma seqüência de vídeos dos sinais que são sincronizados com a trilha de legendas para gerar um vídeo (trilha) em LIBRAS. Por fim, um componente de Distribuição, embarca esse vídeo de LIBRAS no conteúdo multimídia original, incluindo uma mídia extra de LIBRAS, e tornando-o, portanto, acessível para surdos.

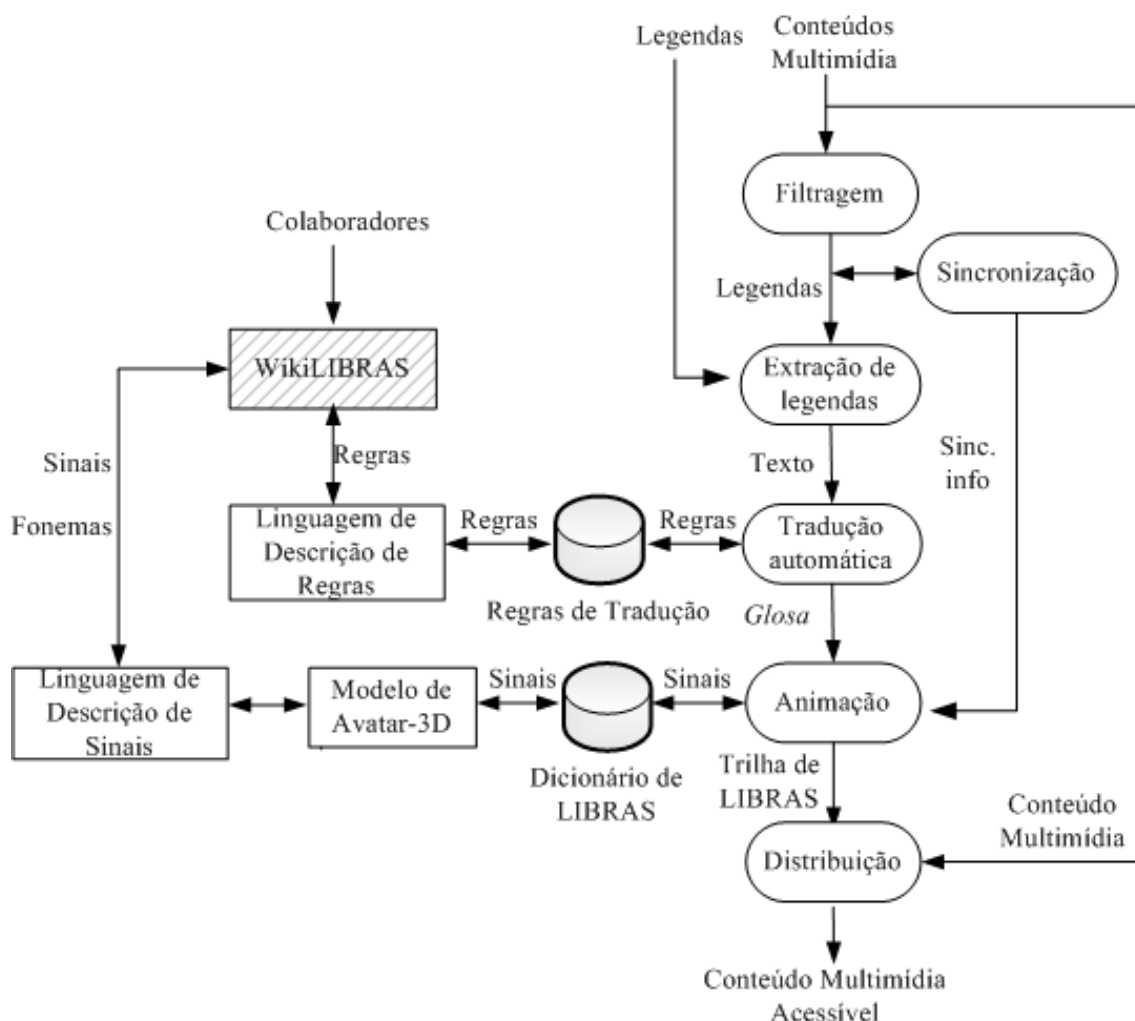
A sincronização entre o conteúdo multimídia original e o vídeo de LIBRAS é realizado com base no modelo de sincronização baseado no eixo do tempo [Blakowski & Steinmetz 1996]. Esse modelo define pontos de sincronização que são inseridos no conteúdo usando etiquetas de tempo (*timestamps*) baseadas em um relógio global. Na solução proposta, o relógio global é o relógio de referência da trilha de legenda. Esse relógio é extraído e utilizado para gerar as etiquetas de tempo de apresentação dos sinais do vídeo (trilha) de LIBRAS. Maiores detalhes sobre os componentes da solução são apresentados nas próximas seções.

4.1 Filtragem e Extração de Legendas

Os componentes de Filtragem e Extração de Legendas são os componentes responsáveis, respectivamente, por filtrar os pacotes de legenda dentro dos conteúdos multimídia e por extrair textos e informações de sincronização das legendas desses conteúdos. Opcionalmente, as legendas podem ser carregadas separadamente do conteúdo multimídia, sendo submetidas diretamente ao componente de Extração de Legendas.

A implementação desses componentes é dependente de plataforma. Na TV Digital,

Figura 4.1: Visão esquemática da solução proposta



por exemplo, esses componentes podem ser implementados com base nas definições das especificações MPEG-2 Sistemas [ISO/IEC 1996] e ABNT NBR 15606-1 [ABNT 2007], respectivamente. A especificação MPEG-2 Sistemas define como extrair os diversos fluxos elementares (como, por exemplo, áudio, vídeo, *closed caption*, aplicações interativas etc.) em um fluxo de transporte (Filtragem), enquanto que a ABNT NBR 15606-1 define, dentre outras coisas, como extrair textos a partir de um fluxo elementar de *closed caption* (Extração de Legendas). Na plataforma de Cinema Digital, esses componentes podem ser implementados com base na especificação DCP (*Digital Cinema Package*) [Digital Cinema Initiatives 2008], que define como os fluxos de legendas são codificados em Cinema Digital. No Capítulo 5, será discutida a implementação desses componentes nas

plataformas de TV Digital, Cinema Digital e Web.

As informações extraídas por esses componentes são sentenças em Português e informações de sincronização que são encaminhados para os componentes de Tradução Automática e Sincronização, respectivamente. As sentenças são encaminhadas para o componente de Tradução para serem traduzidas para LIBRAS e as informações de sincronização são utilizadas pelo componente de Sincronização para gerar os pontos de sincronização do vídeo de LIBRAS resultante do processo de tradução para LIBRAS.

4.2 Tradução Automática

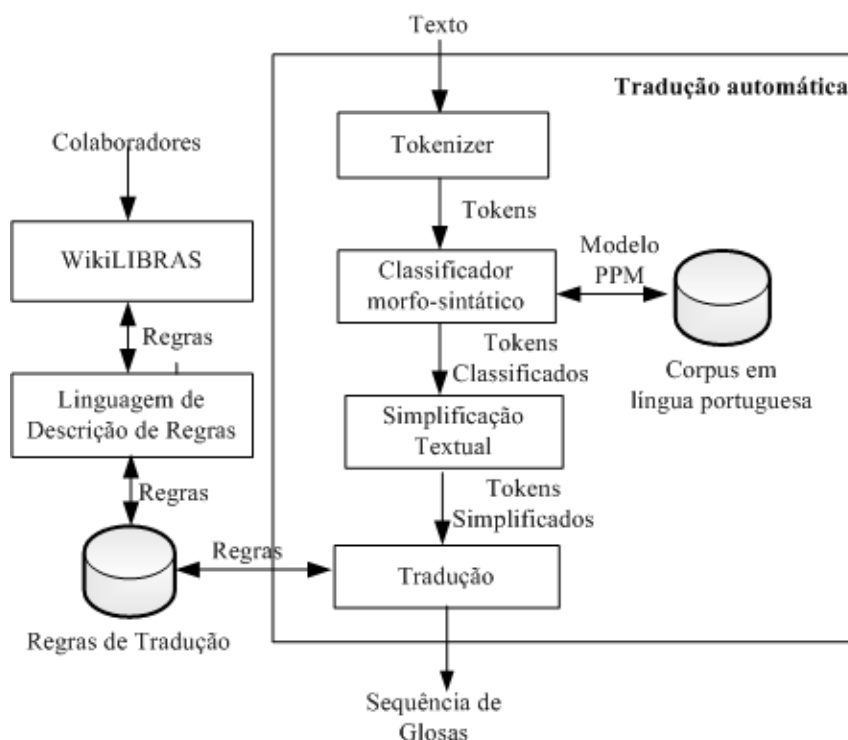
O componente de Tradução Automática converte uma representação textual em língua portuguesa em uma representação textual (seqüência de glosas) em LIBRAS. Conforme mencionado anteriormente, essa estratégia de tradução foi projetada para traduzir conteúdos de forma eficiente (isto é, consumindo pouco tempo) e para domínios gerais. Para isso, ela combina métodos de compressão estatística utilizados para classificar os *tokens* (palavras) de entrada, estratégias de simplificação textual para reduzir a complexidade do texto de entrada e um conjunto de regras morfológicas e sintáticas definido por especialistas. Uma visão esquemática desse componente é ilustrado na Figura 4.2.

De acordo com a Figura 4.2, inicialmente o texto em língua portuguesa é quebrado em uma seqüência de palavras ou *tokens*. Em seguida, esses *tokens* são classificados em categorias morfo-sintáticas. Para essa tarefa de classificação é utilizado o algoritmo PPM-C [Moffat 1990], uma variante do algoritmo *Prediction by Partial Matching* (PPM) [Cleary & Witten 1984]. O PPM é um método de compressão de dados estatístico baseado em modelos de Markov de ordem N e foi escolhido devido a sua habilidade de construir modelos estatísticos com boa precisão [Batista & Meira 2004] e também devido ao seu uso em outros problemas de classificação [Bratko et al. 2006][Mahoui et al. 2008][Medeiros et al. 2011].

O PPM constrói um modelo estatístico a partir de um conjunto de dados de entrada (conjunto de treinamento) e utiliza esse modelo para armazenar a freqüência das diferentes seqüências de elementos encontrados. Após a construção desse modelo, o próximo elemento pode ser predito a partir dos N elementos anteriores. A variante PPM-C é mais eficiente que a implementação original em termos de tempo de execução e espaço de armazenamento utilizado em detrimento de uma taxa de compressão inferior [Moffat 1990].

O Classificador morfológico-sintático modela classes morfológicas e sintáticas como elementos no algoritmo PPM-C. Esse modelo armazena seqüências de classes morfológicas e sintáticas obtidas a partir de um corpus de textos classificados morfológica e

Figura 4.2: Visão esquemática do componente de Tradução Automática



sintaticamente em língua portuguesa. Quando uma sentença é recebida para classificação, a classe morfo-sintática com maior probabilidade para cada *token* é selecionada com base no modelo PPM associado.

Após a classificação desses *tokens*, uma estratégia de simplificação/adaptação textual é aplicada para reduzir a complexidade do texto de entrada. Inicialmente, o texto é simplificado através da remoção de algumas classes de *tokens* como, por exemplo, artigos que não são definidos em LIBRAS. Em seguida, alguns *tokens* são substituídos para adaptar o significado da sentença reescrita para LIBRAS, uma vez o vocabulário de LIBRAS é menor do que o vocabulário da língua portuguesa (substituição léxica) [Santos et al. 2009]. Por exemplo, as palavras "casa", "lar", "habitação" em língua portuguesa são representados pelo mesmo sinal em LIBRAS, o sinal "CASA". Além disso, enquanto os verbos em língua portuguesa possuem um alto grau de inflexão, os verbos em LIBRAS não sofrem inflexão. Dessa forma, os verbos em língua portuguesa são substituídos por verbos no infinitivo em glosa. Para realizar essas substituições, um dicionário de sinônimos de Português para LIBRAS projetado por especialistas em LIBRAS foi utilizado. Finalmente, nomes próprios e termos técnicos são representados em LIBRAS de forma soletrada, isto é, através de configurações de mão que representam as letras do *token* (ou palavra). Dessa

forma, uma substituição datilológica também é aplicada para substituir nomes próprios e termos técnicos.

Por fim, um conjunto de regras de tradução é aplicado para traduzir esses *tokens* restantes para uma representação em glosa. Essas regras de tradução são carregadas a partir de uma base de dados de Regras de Tradução e são descritas usando uma linguagem formal denominada **Linguagem de Descrição de Regras de Tradução**, que será apresentada em detalhes na Seção 4.5.2.

4.3 Animação e Sincronização

O componente de Animação é responsável por converter a seqüência de glosas gerada pelo componente de Tradução Automática em um vídeo em LIBRAS. Para realizar essa tarefa, ele utiliza um Dicionário de LIBRAS que contém uma representação visual (por exemplo, uma animação ou um arquivo de vídeo) para cada sinal. Mais especificamente, nesse dicionário, cada sinal é representado por uma animação ou arquivo de vídeo e possui um código (a glosa) associado com essa representação.

Dessa forma, o Dicionário de LIBRAS pode ser definido formalmente como um conjunto de tuplas t no seguinte formato:

$$t = \langle g, v \rangle, \quad (4.1)$$

onde:

- g é a glosa (ou código) do sinal;
- v é a representação visual do sinal;

Como cada sinal possui um código fixo, a representação visual do sinal pode ser customizada (personalizada). Dessa forma, essa representação pode ser uma animação (ou vídeo) gerada(o) a partir de um agente animado virtual (um avatar) ou um vídeo gravado com um intérprete de LIBRAS. Outra característica importante é que as especificidades regionais da LIBRAS também podem ser respeitadas, uma vez que diferentes dicionários podem ser utilizados na solução dependendo da região onde ela é aplicada.

A utilização de vídeos gravados com intérpretes, no entanto, possui alguns problemas. Um dos principais problemas é que para construir sentenças a partir de vídeos (dos sinais) isolados, é necessário que os vídeos de todos os sinais tenham sido gravados com o mesmo intérprete nas mesmas condições (isto é, mesma roupa, iluminação, distância para câmera,

dentre outros). Caso contrário, a transição entre os sinais consecutivos não será suave [Elliott et al. 2004].

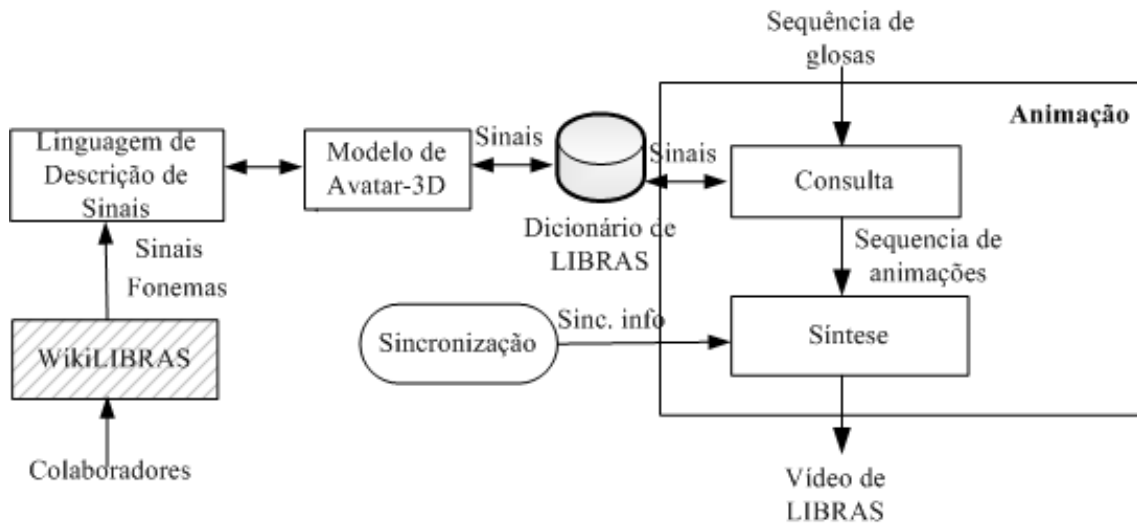
Outro problema está relacionado a atualização do dicionário. Uma vez que LIBRAS é uma língua viva e novos sinais podem surgir espontaneamente, seria necessário gravar novos vídeos para esses novos sinais com o mesmo intérprete nas mesmas condições dos sinais anteriores. Somado-se a disso, a geração desse dicionário demanda, em geral, muito tempo e possui um custo operacional alto.

Em consequência disso, na solução proposta, os sinais do Dicionário de LIBRAS são representados através de um agente animado virtual-3D (um avatar-3D). Com a utilização de avatares é possível gerar todos os sinais sob as mesmas condições e atualizar o Dicionário à medida que novos sinais forem surgindo. Além disso, na solução proposta, os sinais do Dicionário de LIBRAS podem ser desenvolvidos de forma produtiva e melhorados sempre que necessário, através da utilização de ferramentas de colaboração e computação humana (de forma similar a geração de regras de tradução). Para isso, uma **Linguagem de Descrição de Sinais** foi desenvolvida, permitindo que surdos e especialistas em LIBRAS descrevam sinais na ferramenta WikiLIBRAS. A partir dessa descrição, os sinais podem ser renderizados pela WikiLIBRAS com base no **modelo do avatar-3D** proposto. A ferramenta WikiLIBRAS, a Linguagem de Descrição de Sinais e o modelo do avatar-3D serão apresentados na Seção 4.5.

Na Figura 4.3 é ilustrada uma visão esquemática do componente de Animação. De acordo com a Figura 4.3, o componente recebe, inicialmente, uma seqüência de glosas em LIBRAS. A partir dessa seqüência de glosas, uma consulta no Dicionário de LIBRAS é aplicada, associando e recuperando um vídeo para cada glosa (sinal) da sentença. Caso não exista um vídeo para algum dos sinais da sentença no Dicionário de LIBRAS (isto é, não exista uma entrada correspondente para aquele sinal - glosa - no Dicionário de LIBRAS), um vídeo é gerado para aquele sinal a partir da soletração da sua glosa. Essa estratégia é utilizada para evitar lacunas na representação das sentenças em LIBRAS e é a mesma estratégia utilizada pelos surdos e pelos intérpretes de LIBRAS para representar palavras ou termos que não possuem sinais próprios, como, por exemplo, os nomes próprios e os termos técnicos. Por exemplo, em LIBRAS, o nome próprio "TIAGO" é representado como "T-I-A-G-O".

Após recuperar os vídeos dos sinais, o componente de Animação aplica uma estratégia de síntese para combinar esses vídeos, gerando um fluxo único de vídeo de LIBRAS. Essa estratégia consiste em concatenar (combinar) os vídeos com base nas etiquetas de tempo (pontos de sincronização) geradas pelo componente de Sincronização. Para que as transições entre os sinais (vídeos dos sinais) fossem suaves, uma configuração neutra

Figura 4.3: Visão esquemática do componente de Animação



(isto é, posição do avatar, cor de *background*, brilho, etc.) foi definida no início e no fim do vídeo de cada sinal e durante os intervalos de silêncio. No Capítulo 6, será descrito um processo de experimentação conduzido com usuários surdos para avaliar, dentre outros, a naturalidade e inteligibilidade dos vídeos de LIBRAS produzidos pela solução utilizando essa estratégia de suavização.

Conforme mencionado anteriormente, a estratégia de sincronização utilizada pela solução proposta é baseada no modelo de sincronização baseado no eixo do tempo [Blakowski & Steinmetz 1996]), onde o relógio global do conteúdo multimídia é utilizado como referência para gerar as etiquetas de tempo de apresentação (*Presentation Timestamps* - PTS) que servem como pontos de sincronização para a trilha de LIBRAS. Considerando que rel_{mm} e $pulso_{mm}$ sejam, respectivamente, o relógio de referência do conteúdo multimídia e pulso desse relógio; rel_{leg} e $pulso_{leg}$ sejam, respectivamente, o relógio e o pulso do relógio do fluxo de legenda; e t_{leg} seja a marcação de tempo do bloco de legenda, o componente de Sincronização define o valor de PTS para o trecho do vídeo de LIBRAS (PTS_{Libras}) correspondente de acordo com a seguinte equação:

$$PTS_{Libras} = rel_{mm} + \frac{(t_{leg} - rel_{leg}) * pulso_{mm}}{pulso_{leg}}, \quad (4.2)$$

De acordo com a Equação 4.2, caso o fluxo de legenda esteja embutido no próprio conteúdo multimídia, o relógio e o pulso do fluxo de legenda coincidirão com o relógio e o pulso do conteúdo multimídia e, com isso, o valor de PTS (etiqueta de tempo)

para o trecho do fluxo de LIBRAS será o mesmo do bloco de legenda correspondente (t_{leg}), sincronizando as duas mídias (legenda e LIBRAS). Caso contrário, os relógios dos dois fluxos podem ser diferentes, e, nesse caso, os valores de PTS do fluxo de LIBRAS precisam ser adaptados de acordo com a Equação 4.2.

Esses valores de PTS gerados pelo componente de Sincronização são então encaminhados para o componente de Animação que insere essas etiquetas (pontos de sincronização) no fluxo de vídeo (trilha) de LIBRAS, sincronizando a trilha de LIBRAS com a trilha de legenda.

4.4 Distribuição

Após a geração do vídeo de LIBRAS, a solução acopla esse vídeo no conteúdo multimídia original de acordo com as características da plataforma, adicionando o conteúdo de LIBRAS naquele contexto. Essa tarefa é realizada pelo componente de Distribuição e pode ser feita de várias formas. São elas:

1. Mixar o vídeo de LIBRAS no conteúdo original: Nesse caso, os quadros do vídeo de LIBRAS são apresentados em uma janela sobreposta aos quadros do conteúdo multimídia original. Essa estratégia é interessante porque torna a exibição do conteúdo (trilha) de LIBRAS independente do *player* de vídeo. Contudo, uma limitação dessa abordagem é que após ser aplicado, não é mais possível desabilitar ou remover o vídeo de LIBRAS.
2. Multiplexar o vídeo de LIBRAS no conteúdo original: Nesse caso, o vídeo de LIBRAS é codificado como fluxo de vídeo independente e separado do conteúdo multimídia original mas é encapsulado (envelopado) junto com este em um fluxo único de transporte como, por exemplo, no protocolo de transporte MPEG-2 *Transport Stream* (MPEG-2 TS) [ISO/IEC 1996]. Dessa forma, pode-se ter um fluxo único de transporte contendo as duas trilhas de vídeo. Essa abordagem, no entanto, torna o vídeo acessível dependente do *player* de vídeo que precisa ser capaz de interpretar o protocolo de transporte e tocar os dois vídeos ao mesmo tempo. Por outro lado, é possível habilitar, desabilitar, reposicionar ou redimensionar o vídeo (trilha) de LIBRAS.
3. Transmitir o vídeo de LIBRAS para outros dispositivos: Nesse caso, o vídeo de LIBRAS é transmitido como um fluxo de vídeo para ser apresentado em outro *display* como, por exemplo, em um dispositivo móvel. Essa abordagem é interessante em situações onde não é interessante exibir o vídeo de LIBRAS diretamente sobre

o conteúdo multimídia original, como por exemplo, nos cinemas, onde o ambiente de exibição do conteúdo multimídia é compartilhado e a janela de LIBRAS poderia incomodar os usuários que não são surdos. Nesse caso, seria possível transmitir o vídeo de LIBRAS para ser exibido de forma independente num *display* específico do usuário (como, por exemplo, no seu *smartphone* ou *tablet*). Considerando que, de acordo com a norma ABNT NBR 15290 [ABNT 2005], são tolerados atrasos de até quatro segundos em sistemas de *closed caption* ao vivo, é possível admitir nesse contexto, portanto, atrasos e variações de retardo de transmissão de até quatro segundos.

4.5 Ferramenta de Computação Humana

Nessa seção, será apresentada a estratégia utilizada para o desenvolvimento eficiente das construções lingüísticas (regras de tradução e sinais) da solução proposta. Essa estratégia é composta por uma ferramenta de colaboração, denominada WikiLIBRAS, que controla a geração dessas construções/conteúdos, por linguagens formais para descrição das regras de tradução e sinais e pelo modelo de um agente animado virtual-3D (um avatar-3D) utilizado para representar os sinais de LIBRAS na solução. Na Seção 4.5.1 será apresentada a arquitetura da ferramenta WikiLIBRAS. Nas Seções 4.5.2, 4.5.3 e 4.5.4 são apresentadas as linguagens de descrição de regras de tradução, de sinais e o modelo do agente animado virtual-3D, respectivamente.

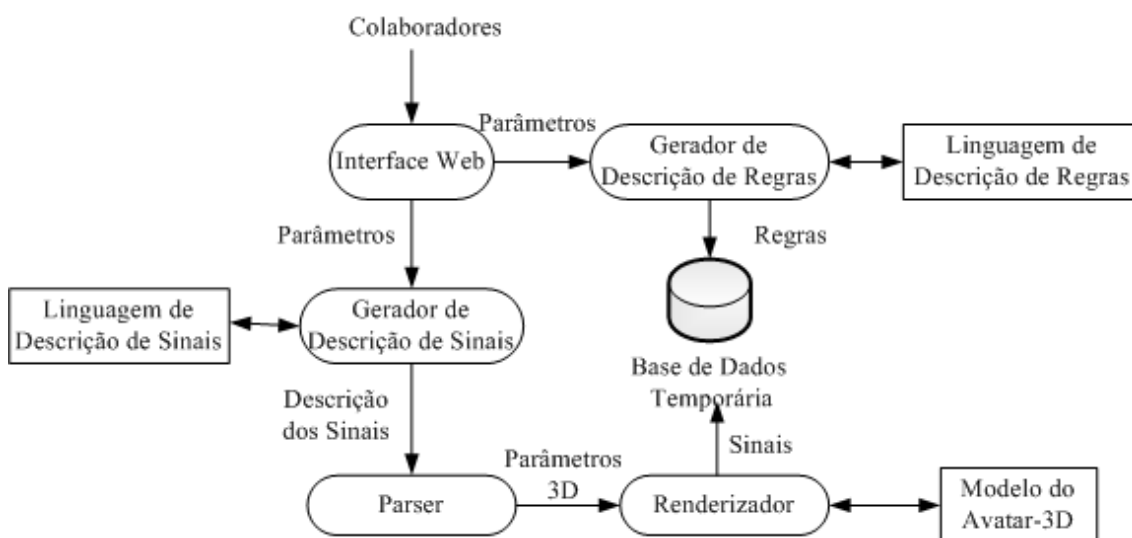
4.5.1 WikiLIBRAS

Conforme mencionado anteriormente, os componentes de Tradução Automática e Animação dependem, respectivamente, de bases de dados compostas por regras de tradução e sinais em LIBRAS (as construções lingüísticas da solução). Considerando que o desenvolvimento dessas construções é uma tarefa que demanda o conhecimento de especialistas e que, em geral, consome muito tempo, uma ferramenta, denominada WikiLIBRAS, foi desenvolvida para auxiliar no processo de desenvolvimento dessas construções/conteúdos. A idéia do WikiLIBRAS é permitir que colaboradores possam participar do processo de desenvolvimento dessas construções através da adição de novas construções ou da edição das construções existentes, tornando o seu desenvolvimento mais produtivo.

Além disso, a WikiLIBRAS também propicia a capacidade de melhorar a qualidade dos conteúdos gerados pela solução sempre que necessário. Por exemplo, uma vez que

os usuários avaliem que a qualidade de um determinado sinal ou uma determinada regra de tradução não estão produzindo resultados apropriados, eles podem editar esse sinal ou regra, melhorando a qualidade dos conteúdos gerados pela solução. Na Figura 4.4, é apresentada a arquitetura do WikiLIBRAS.

Figura 4.4: Visão esquemática da ferramenta WikiLIBRAS



De acordo com a Figura 4.4, inicialmente, os colaboradores acessam o ambiente colaborativo através de uma interface Web. A partir do acesso a essa interface, eles podem configurar novas regras e sinais ou buscar novos sinais ou regras já existentes para edição. Quando o usuário deseja adicionar uma nova regra de tradução, um módulo Gerador de Descrição de Regras converte a interação do usuário (configuração de parâmetros da regra de tradução) numa representação formal, de acordo com a Linguagem de Descrição de Regras de Tradução (que será apresentada na Seção 4.5.2). Essa representação é então armazenada numa base de dados temporária para ser avaliada por especialistas em LIBRAS, isto é, uma etapa de supervisão é aplicada antes da regra entrar na base de dados da solução.

Esta etapa de supervisão evita que regras de tradução incorretas sejam adicionadas na base de dados de Regras de Tradução por usuários não-especialistas ou mal intencionados. Após a sua aprovação, a regra é adicionada a base de dados de Regras de Tradução e passa a compor a solução. Adicionalmente, os usuários também podem pesquisar as regras de tradução existentes e editá-las. Assim como acontece quando uma nova regra é adicionada, a edição também será supervisionada antes de ser aplicada na base de dados.

De forma análoga a geração de novas regras, quando o usuário configura um novo sinal, um módulo Gerador de Descrição de Sinais converte as interações dos usuários em uma representação formal de acordo com a Linguagem de Descrição de Sinais (que será apresentada na Seção 4.5.3). Em seguida, essa representação é convertida pelo módulo *Parser* para um conjunto de parâmetros baseada no modelo de um agente animado virtual (um avatar-3D, que será apresentado na Seção 4.5.4) e uma vídeo do sinal é renderizado pelo módulo Renderizador a partir desses parâmetros. Esse vídeo do sinal é então devolvida para o usuário que pode avaliar se ela foi gerada corretamente. De forma análoga, ao processo de geração de novas regras, os novos sinais gerados na ferramenta também passam por um processo de supervisão antes de serem incluídas no Dicionário de LIBRAS.

4.5.2 Linguagem de Descrição de Regras de Tradução

Para representar as regras que compõem a base de dados de Regras de Tradução, uma Linguagem de Descrição de Regras de Tradução (LDRT) foi definida. Essa linguagem permite que os parâmetros configurados pelos especialistas na WikiLIBRAS sejam representadas de maneira formal para serem aplicados pelo componente de Tradução Automática da solução proposta.

Nessa linguagem cada regra de tradução é definida formalmente como uma tupla r no seguinte formato:

$$r = \langle e_1, e_2, \dots, e_c \rangle, \quad (4.3)$$

Onde e_1, e_2, \dots, e_n é um conjunto de elementos lingüísticos ordenados de acordo com a sentença de entrada e c é o número de elementos. Esses elementos lingüísticos e_i são formalmente definidos da seguinte forma:

$$e_i = \langle ms_{class}, n_{pos}, n_{prop} \rangle, i = 1, 2, \dots, c \quad (4.4)$$

Onde ms_{class} representa a classe morfológica e sintática do elemento lingüístico, n_{pos} representa a nova posição desse elemento após a aplicação da regra, com o valor -1"indicando que o elemento deve ser removido da sentença, e n_{prop} é um campo opcional que indica possíveis mudanças no elemento lingüístico (por exemplo, todo verbo em LIBRAS deve estar no infinitivo). A lista de classes morfo-sintáticas e seus rótulos (valores do campo ms_{class}) são ilustradas na Tabela 4.1.

Baseado nessas definições, uma representação XML foi especificada para representar os atributos das regras de tradução. Cada regra possui um campo contador que representa

Tabela 4.1: Classes morfo-sintática da Linguagem de Descrição de Regras de Tradução

Classe morfológica	Rótulo	Classe Sintática	Rótulo
substantivos	sub	sujeito	SUBJ
adjetivos	adj	verbo transitivo	VERT
nomes próprios	prop	verbo intransitivo	VERI
advérbios	adv	predicativo do sujeito	SC
verbos	ver	predicativo do objeto	OC
verbos flexionados	ver-f	objeto (direto ou indireto)	OBJ
verbos no gerúndio	ver-g	objeto direto	OBJI
verbos no particípio	ver-p	objeto indireto	OBJI
verbos no infinitivo	ver-i	complemento nominal	COMP NOM
artigos	art	agente da passiva	PASS
pronomes	pron	adjunto adnominal	ADJ NOM
preposições	pre	adjunto adverbial	ADJ VER
interjeições	intj	aposto	APP
conjunções	conj	tópico	TOP
conjunções subordinativas	conj-s		
conjunções coordenativas	conj-c		
prefixos	ec		

o número de elementos lingüísticos. Para cada elemento, existe um campo *titulo* que representa a classe morfo-sintática (conforme os rótulos definidos na Tabela 4.1), um campo *novapos* (atributo n_{pos}) que indica a nova posição do elemento após a aplicação da regra, e um campo opcional *novaprop* que representa o atributo n_{prop} e indica possíveis mudanças nos elementos. Cada regra possui também um *flag* ativa que indica se a regra deve ou não ser aplicada pela solução, facilitando o processo de testes com as regras de tradução. Para que a regra possa ser aplicada, os elementos do texto original devem aparecer na mesma ordem definida na regra.

A Figura 4.5 ilustra um exemplo de uma representação XML de uma regra de tradução. Segundo especialistas em LIBRAS, essa regra é a regra de tradução mais comum para tradução de português para LIBRAS (regra de topicalização), e indica que sempre que for encontrada uma sentença com os *tokens* de entrada na ordem "Sujeito+Verbo+Objeto", eles devem ser reposicionados de forma que o Objeto venha primeiro, seguida pelo Sujeito e depois pelo Verbo no texto traduzido.

Figura 4.5: Exemplo de representação da regra de topicalização: uma sentença na ordem sujeito+verbo+objeto é traduzida como objeto+sujeito+verbo.

```
<regra>
  <ativa>true</ativa>
  <contador>3</contador>
  <classe>
    <titulo>SUBJ</titulo>
    <novapos>1</novapos>
  </classe>
  <classe>
    <titulo>VERT</titulo>
    <novapos>2</novapos>
    <novaprop>inf</novaprop>
  </classe>
  <classe>
    <titulo>OBJ</titulo>
    <novapos>0</novapos>
  </classe>
</regra>
```

4.5.3 Linguagem de Descrição de Sinais

De forma análoga as regras de tradução, uma linguagem também foi desenvolvida para representar formalmente os sinais que irão compor o Dicionário de LIBRAS. Essa linguagem, denominada Linguagem de Descrição de Sinais (LDS) permite que os parâmetros que compõem um sinal sejam descritos, e a partir desses parâmetros, vídeos dos sinais podem ser geradas com base no modelo de um agente animado virtual-3D (que será na Seção 4.5.4).

Nessa linguagem, cada sinal em LIBRAS é definido através de um conjunto de movimentos, onde cada movimento possui uma configuração inicial e final das mãos, braços e face, um tipo de trajetória (por exemplo, retilíneo, circular, semicircular etc.), uma direção (por exemplo, de dentro pra fora, da direita pra esquerda, etc.), além de *flags* que indicam quais mãos são usadas no movimento (direita, esquerda ou ambas), entre outros.

Formalmente, um sinal s é definido da seguinte forma:

$$s = \langle gl, mov_1, mov_2, \dots, mov_n \rangle, \quad (4.5)$$

$$mov_i = \langle cf_{ini}, cf_{fin}, traj, dir, me_f, md_f \rangle, i = 1, 2, \dots, n, \quad (4.6)$$

$$cf_t = \langle cfm_e, cfm_d, or_e, or_d, pa_e, pa_d, ef \rangle, t = ini, fin, \quad (4.7)$$

$$or_m = \langle or_palm, sent_palm, sent_ded \rangle, m = e, d, \quad (4.8)$$

$$pa_m = \langle subd, loc \rangle, m = e, d, \quad (4.9)$$

Onde *gl* é a glosa do sinal e $mov_1, mov_2, \dots, mov_n$ são o conjunto de movimentos do sinal. Os parâmetros cf_{ini}, cf_{fin} representam a configuração inicial e final de cada movimento (mov_i), respectivamente; *traj* e *dir* representam o tipo de trajetória e a direção de cada movimento, respectivamente e me_f e md_f indicam, respectivamente, se as mãos esquerda e direita são utilizadas no movimento. cfm, or, pa e *ef* representam a configuração de mão, a orientação da palma da mão (por exemplo, para cima, para baixo, para frente, para trás etc), o ponto de articulação e a expressão facial de cada configuração. Os índices *e* e *d* dos fonemas *hs, or, loc* referem-se à mão direita ou esquerda, respectivamente. Por fim, os parâmetros *or_palm, sent_palm* e *sent_ded*, representam, respectivamente, o plano de orientação da palma da mão (paralelo ao corpo ou paralelo ao solo), sentido adotado pela palma da mão e o sentido da ponta dos dedos, enquanto que os parâmetros *loc* e *subd* representam, respectivamente, a localização no corpo (cabeça, corpo ou no espaço neutro) e subdivisão dentro dessa localização.

A partir dessa formalização, foi definida uma representação XML para representar esses parâmetros e, conseqüentemente, descrever sinais. As Figuras 4.6 e 4.7 ilustram exemplos das representações XML dos sinais LÁBIO e CORRETO em LIBRAS, respectivamente.

De acordo com a Figura 4.6, o atributo tipo representa o tipo de trajetória do movimento. Os atributos *mao-usada* e *flag-repeticao* representam, respectivamente, às *flags* que indicam quais são as mãos utilizadas no movimento (esquerda, direita ou ambas) e se há ou não repetição do movimento descrito. Os atributos *sentido* e *tamanho-raio* são parâmetro exclusivos do movimentos circulares e representam, o sentido do movimento (horário ou anti-horário, por exemplo), e o tamanho do raio (pequeno, médio ou grande) desse movimento circular. O campo *config-mao* indica qual é a configuração de mão utilizada no movimento, que é representada por um valor inteiro entre 1 a 60, de acordo com as opções de configuração da mão apresentadas na Figura 2.1. A orientação da palma da mão (campo *config-palma*) possui os atributos *orientação*, *sentido* e *sentido-dedos* que se referem, respectivamente, ao referencial adotado da mão em relação ao corpo, o sentido adotado pela palma da mão e o sentido da ponta dos dedos. Por

Figura 4.6: Exemplo de representação do sinal LÁBIO

```

<sinal>
  <nome>LABIO </nome>
  <movimento tipo="circular"tamanho-raio="pequeno"sentido="horario"
  flag-repeticao="sem-repeticao"mao-usada="direita">
    <config>
      <config-mao> 14 </config-mao>
      <config-palma> orientacao="paralela-ao-corpo"
      sentido-palma="para-tras"sentido-dedos="para-cima"
      </config-palma>
    </config>
    <ponto-articulacao local="cabeca"subdivisao="boca"/>
    <expressao-facial="neutra">
  </movimento>
</sinal>

```

Figura 4.7: Exemplo de representação do sinal CORRETO.

```

<sinal>
  <nome>CORRETO </nome>
  <movimento tipo="retilineo"flag-repeticao="sem-repeticao"mao-usada="direita">
    <config estado="inicial">
      <config-mao> 17 </config-mao>
      <config-palma> orientacao="paralela-ao-solo"sentido-palma="para-dentro"
      sentido-dedos="para-frente"
      </config-palma>
    </config>
    <ponto-articulacao estado="inicial"local="espaco-neutro"
    subdivisao="busto-direito-mediana">
    <ponto-articulacao estado="final"local="espaco-neutro"
    subdivisao="cintura-direita-mediana">
    <expressao-facial="neutra">
  </movimento>
</sinal>

```

fim, o campo *expressao-facial* representa a expressão facial (expressão não-manual) do sinal representado. Os valores definidos para os principais parâmetros desta linguagem são apresentados no Apêndice C.

Dessa forma, na Figura 4.6, o sinal LÁBIO foi definido com a mão direita (com a configuração de mão 14 - ver Figura 2.1) realizando um movimento circular ao redor da boca. As configuração da mão, orientação da palma da mão e o ponto de articulação não mudam durante o movimento e, portanto, as configurações iniciais e finais são iguais. Na Figura 4.7, o sinal CORRETO foi definido apenas com a mão direita (com a configuração de mão 17 - ver Figura 2.1) e com ponto de articulação inicial e final diferentes. Mais especificamente, no sinal, um movimento linear (tipo de trajetória do movimento) é realizado do busto direito (ponto de articulação inicial) para a cintura direita (ponto de articulação final).

4.5.4 Modelo de um Agente Animado Virtual-3D

Para representar os sinais descritos pela Linguagem de Descrição de Sinais na solução proposta, um agente animado virtual 3D (um avatar-3D) foi modelado e implementado. Esse avatar-3D foi modelado no *software* Blender¹ com uma armadura composta por 82 ossos, distribuídos da seguinte forma:

- 15 ossos em cada uma das mãos para configurar a posição dos dedos;
- 23 ossos para configurar os elementos faciais;
- 22 ossos para configurar os movimentos de braço e de corpo;
- 7 ossos auxiliares que não deformam a malha do avatar-3D diretamente.

Dessa forma, para configurar, por exemplo, os movimentos dos dedos, é necessário definir os parâmetros de localização e rotação de cada um dos 15 ossos da mão. O mesmo deve ser feito para configurar os ossos da face do avatar-3D. Os movimentos do braço são realizados através da movimentação de apenas 2 ossos. O primeiro deles está localizado no pulso do avatar-3D e o segundo é um osso auxiliar que controla a deformação do cotovelo e antebraço.

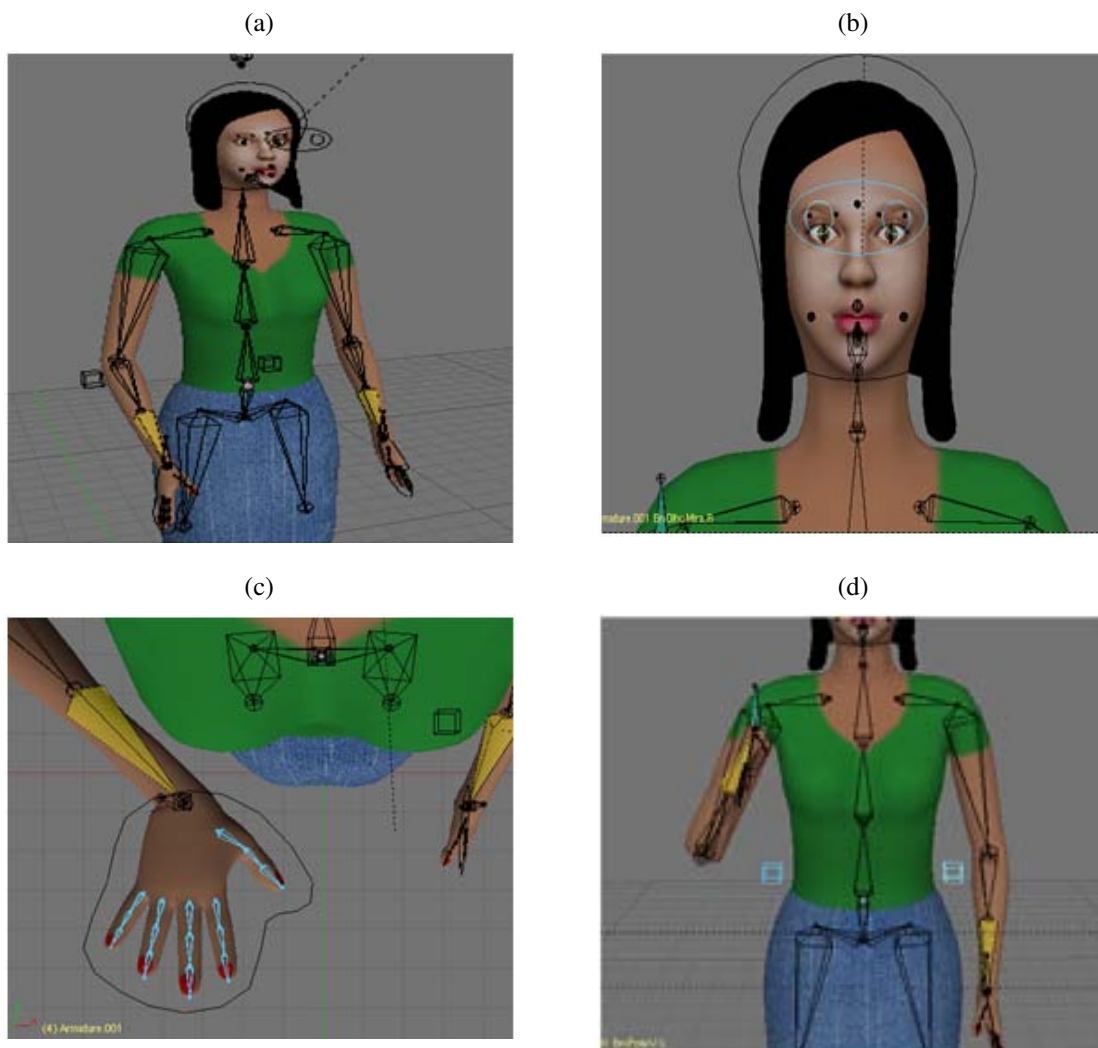
Para combinar a deformação entre ossos relacionados foi utilizada cinemática inversa (*inverse kinematics* - ik). De acordo com a Adobe (2012b), a cinemática inversa (IK) é uma forma de animar objetos usando ossos encadeados em armaduras lineares ou ramificadas em relacionamentos pai-filho. Quando um osso se movimenta, os ossos conectados

¹www.blender.org/

se movem em relação a ele. Dessa forma, se houver, por exemplo, um movimento no osso do pulso, ele irá espalhar para os ossos do braço e do antebraço.

O modelo do avatar-3D é ilustrado na Figura 4.8. As Figuras 4.8b, 4.8c and 4.8d ilustram esse modelo com ênfase nos ossos da face, das mãos e do corpo, respectivamente.

Figura 4.8: (a) Modelo do avatar-3D. Ênfase nos ossos da (b) face, (c) das mãos e (d) do corpo.



4.6 Considerações

Nesse capítulo a arquitetura da solução proposta e suas principais características foram apresentadas. Os principais componentes Extração de Legendas, Tradução Automática, Animação e Distribuição) também foram apresentados de forma detalhada juntamente

com as estratégias desenvolvidas para o desenvolvimento das suas construções lingüísticas (WikiLIBRAS, Linguagem de Descrição de Regras e Sinais e Modelo do Avatar-3D).

No próximo capítulo serão apresentadas implementações dessa solução para as plataformas de TV Digital, Cinema Digital e Web, conduzidas como cenários de uso para a solução proposta. No Capítulo 6 será apresentado um conjunto de experimentos conduzidos para avaliar a solução com relação a qualidade dos conteúdos gerados, nível de compreensão dos usuários surdos, eficácia e eficiência do WikiLIBRAS na geração das construções lingüísticas, além de testes computacionais para avaliar o desempenho de tradução.

Capítulo 5

Cenários de Uso

Com o objetivo de construir provas de conceito da solução proposta, nesse Capítulo serão apresentadas implementações da solução proposta para TV Digital, Web e Cinema Digital. Na Seção 5.1, será apresentado o LibrasTV, protótipo da solução proposta desenvolvido para o Sistema Brasileiro de TV Digital (SBTV D) e nas Seções 5.2 e 5.3 serão apresentados, respectivamente, o LibrasWeb e o CineLibras, protótipos desenvolvidos para Web e Cinema Digital, respectivamente. Por fim, na Seção 5.4, serão apresentados os detalhes de implementação do WikiLIBRAS.

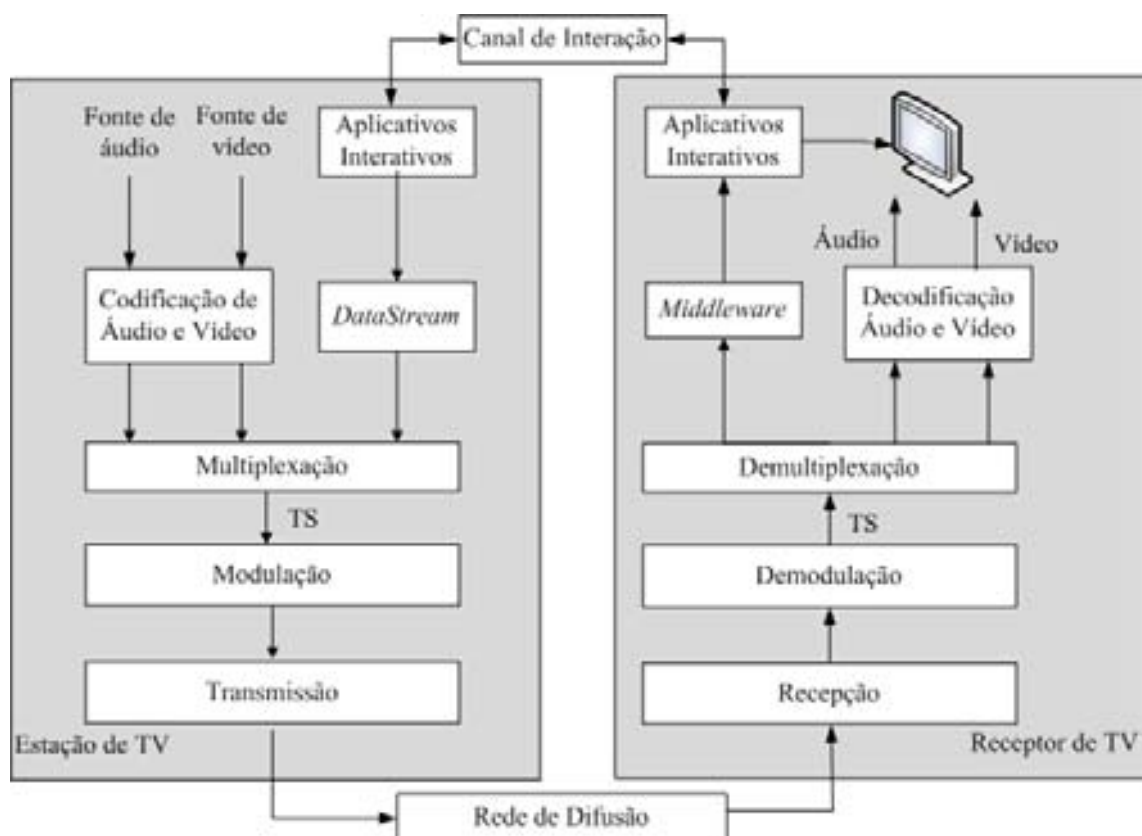
5.1 LibrasTV

De acordo com Souza Filho et al. (2007), a TV Digital Interativa pode ser definida como a fusão da TV tradicional com as tecnologias computacionais. Dessa forma, o impacto da mudança da TV Analógica para TV Digital é muito mais significativo do que uma simples troca do sistema de transmissão do analógico para o digital. A TV Digital permite, por exemplo, que aplicações interativas sejam transmitidas juntamente com o áudio e vídeo do programa, tornando possível a expansão das funcionalidades do sistema [Soares & Barbosa 2005]. Além disso, com a TV digital também é possível transmitir sinais com melhor qualidade, como, por exemplo, vídeos em alta-definição e áudios multi-canais, é possível transmitir vários fluxos de vídeos independentes no mesmo canal físico (multiprogramação), dentre outros.

Um sistema de TV Digital é basicamente um sistema cliente-servidor, onde o servidor é a estação de TV (ou provedor de conteúdo) e os clientes são o ambiente do usuário telespectador (ver Figura 5.1). De acordo com a Figura 5.1, na estação de TV, inicialmente, as fontes analógicas de áudio e vídeo (capturadas de câmeras ou recuperadas de um servidor de vídeo) são entregues aos codificadores digitais, responsáveis por codificar e comprimir os fluxos de áudio e vídeo. Esses fluxos de áudio e vídeo comprimidos são

então multiplexados juntamente com os fluxos de dados em um único fluxo, denominado fluxo de transporte (*Transport Stream - TS*). O fluxo TS é então modulado e transmitido numa rede de difusão (terrestre, cabo ou satélite, por exemplo) para os receptores.

Figura 5.1: Sistema de TV Digital.



No lado do receptor, o sinal é recebido, demodulado e entregue para demultiplexadores que separam os fluxos de áudio, vídeo e dados. Os fluxos de áudio e vídeo são então entregues aos decodificadores, que decodificam e sincronizam esses sinais para apresentação, enquanto que os fluxos de dados são enviados para serem processados pelo *middleware*¹. As aplicações interativas podem também requerer novos dados que podem ser obtidos a partir de um canal de interação (ou canal de retorno).

A integração da solução proposta nos sistemas de TV Digital, denominada LibrasTV, pode ser realizada de diversas formas. Por exemplo, todos os componentes podem ser integrados na estação de TV e o vídeo (trilha) de LIBRAS seria gerado e transmitido

¹O *middleware* é uma camada de software responsável por abstrair as características específicas de cada receptor, permitindo que a mesma aplicação possa ser executada em receptores de diferentes fabricantes.

como um fluxo de vídeo independente para os receptores. Outra opção seria executar todos os componentes no receptor, gerando o vídeo de LIBRAS no receptor, ou então carregando essas informações pelo canal de interação. A solução adotada pelo LibrasTV, no entanto, é baseada na seguinte estratégia (ver Figura 5.2):

- Os componentes de Filtragem, Extração de Legendas e Tradução Automática são agrupados em um módulo denominado "Tradutor de LIBRAS" e integrados na estação de TV (ou provedor de conteúdo). Esse módulo recebe um fluxo de legenda, extrai as sentenças em Português desse fluxo e as traduz para uma seqüência de glosas em LIBRAS. Essa seqüência de glosas é então codificada juntamente com as etiquetas de tempo (informações de sincronização) e encapsulada no fluxo TS com base em um protocolo de codificação que será apresentado na Seção 5.1.1.
- Os componentes de Animação e Distribuição são agrupados e implementados como uma aplicação interativa que será executada nos receptores de TV. Essa aplicação extrai a seqüência de glosas e as informações de sincronismo encapsuladas no fluxo TS, decodifica, sincroniza e apresenta a trilha de LIBRAS com o auxílio do Dicionário de LIBRAS.
- O Dicionário de LIBRAS é carregado pelo canal de retorno ou então armazenado em um dispositivo de memória externa (por exemplo, um dispositivo USB).

Uma das principais vantagens dessa estratégia é a utilização de pouca largura de banda do canal de TV, uma vez que apenas uma seqüência de glosas (texto) codificadas é transmitida no fluxo multiplexado. Outra característica importante é que ela também permite que as diferenças regionais da LIBRAS sejam respeitadas, uma vez que cada usuário pode carregar ou utilizar seu próprio Dicionário de LIBRAS, gerando uma trilha de LIBRAS personalizada de acordo com o Dicionário de LIBRAS utilizado.

Conforme já mencionado, outras possíveis formas de integrar a solução proposta em um sistema de TV Digital seriam:

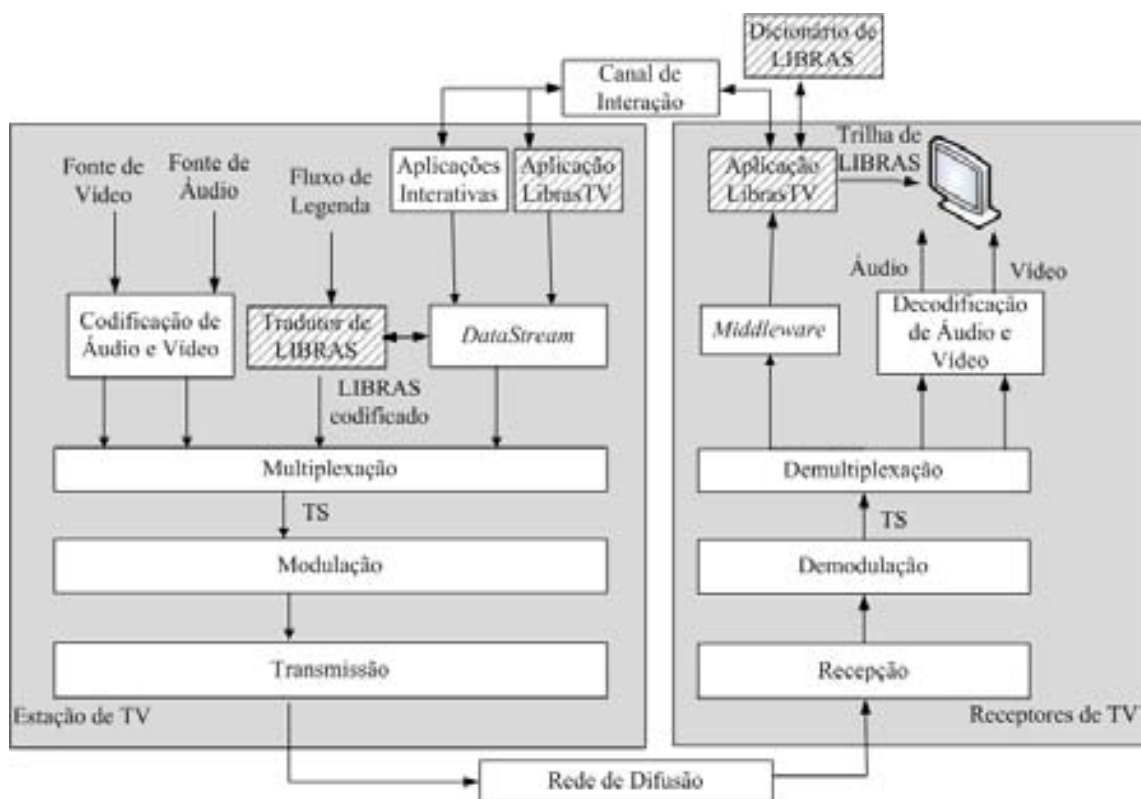
1. Integração de todos os componentes na Estação de TV. Nesse caso, a trilha de LIBRAS seria gerada na estação de TV e transmitida no fluxo MPEG-2 TS como um fluxo de vídeo secundário.
2. Integração de todos os componentes nos receptores de TV. Nesse caso, a trilha de LIBRAS seria gerada e apresentada integralmente no receptor de TV.

Com relação a estratégia proposta, a primeira alternativa utiliza uma maior largura de banda, uma vez que um vídeo secundário é transmitido no fluxo TS, e não preserva as diferenças regionais, uma vez que a mesma trilha de LIBRAS é gerada para todos os

usuários. A segunda solução também possui algumas limitações, uma vez que ela exige uma grande carga de processamento nos receptores de TV para traduzir, sincronizar e apresentar a trilha de LIBRAS.

Esses argumentos justificaram a decisão de integrar a solução proposta conforme apresentado na Figura 5.2. Para implementar essa solução, no entanto, um protocolo de codificação precisa ser definido para inserção da seqüência de glosas e das informações de sincronização no fluxo TS. Esse protocolo é apresentado na próxima seção.

Figura 5.2: Arquitetura geral do protótipo LibrasTV.



5.1.1 Protocolo de Codificação do LibrasTV

O protocolo de codificação do LibrasTV permite que as seqüências de glosas e as informações de sincronização geradas, respectivamente, pelos componentes de Tradução Automática e Sincronização na estação de TV, possam ser encapsuladas no fluxo TS². Ele

²Esse protocolo de codificação está atualmente em discussão no Fórum SBTVD e é um dos candidatos a se tornar o padrão de transmissão de LIBRAS no SBTVD.

é basicamente composto por dois tipos de mensagens: a `LIBRASControlMessage` (LCM), uma mensagem de controle; e a `LIBRASDataMessage` (LDM), uma mensagem de dados.

As mensagens LCM são utilizadas para transmitir periodicamente as configurações iniciais de apresentação da trilha de LIBRAS (posição e tamanho da janela, resolução, dentre outros), enquanto que as mensagens LDM são utilizadas para transmitir a sequência de glosas em LIBRAS. A sintaxe das mensagens LCM e LDM são apresentadas nas Tabelas 5.1 e 5.2, respectivamente.

Tabela 5.1: Sintaxe das mensagens LCM

LCM{	
<code>libras_control_id</code>	8 bits
<code>libras_control_length</code>	16 bits
<code>resolution</code>	8 bits
<code>window_line</code>	16 bits
<code>window_column</code>	16 bits
<code>window_width</code>	16 bits
<code>window_height</code>	16 bits
}	

Tabela 5.2: Sintaxe das mensagens LDM

LDM{	
<code>libras_data_id</code>	8 bits
<code>libras_data_length</code>	16 bits
<code>number_of_signs</code>	16 bits
<code>for (i =0; i < N; i++){</code>	
<code>gloss_bytes_length</code>	8 bits
<code>for (j =0; j < M; j++){</code>	8 bits
<code>gloss_data_bytes</code>	8 bits
}	
}	
}	

De acordo com as Tabelas 5.1 e 5.2, as mensagens LCM e LDM iniciam com um campo de identificação e um campo de comprimento da mensagem (os campos `libras_control_id` e `libras_control_length` para a LCM, e os campos `libras_data_id` e `libras_data_length` para

a LDM). Esses campos são utilizados para identificar o tipo de mensagem (LCM ou LDM) e comprimento da mensagem em bytes, respectivamente.

A mensagem LCM também é composta pelos seguintes campos: `resolution`, `window_line`, `window_column`, `window_width` e `window_height`. O campo `resolution` representa a resolução da camada gráfica utilizada na janela de apresentação da trilha de LIBRAS (por exemplo, 1920 x 1080, 720 x 480, etc.). Os possíveis valores para o campo `resolution` são apresentados na Tabela 5.3. Os campos `window_line` e `window_column` definem as coordenadas da posição inicial da janela de LIBRAS (tomando o canto superior esquerdo como referência) na camada gráfica, enquanto que os campos `window_width` e `window_height` definem a largura e altura da janela inicial.

Tabela 5.3: Valores do campo `resolution`.

Valores	Resolução
0	1920x1080
1	1280x720
2	640x480
3	960x540
4	720x480
5	320x240
6-255	reservado para uso futuro

Nas mensagens LDM, o campo `gloss_data_bytes` transporta as glosas que estão sendo codificadas. Uma vez que esses campos são definidos dentro de um laço de repetição, várias glosas podem ser transmitidas na mesma mensagem. O campo `number_of_signs` especifica o número de glosas que estão codificadas na LDM.

Para encapsular as mensagens LCM e LDM no fluxo MPEG-2 TS, o protocolo de transporte adotado por todos os sistemas de TV atuais [ISO/IEC 1996], uma alternativa é utilizar os eventos definidos na especificação *Digital Storage Media-Command and Control* (DSM-CC) [ISO/IEC 1998], denominados eventos DSM-CC.

Os eventos DSM-CC são transmitidos (encapsulados) em estruturas denominadas Stream Event Descriptors (ver Tabela 5.4), e permitem que pontos de sincronização sejam definidos no nível da aplicação, tornando possível a sincronização da aplicação com outras mídias relacionadas como, por exemplo, o áudio e o vídeo do programa. Esta estrutura é composta basicamente por um campo de identificação (`eventID`), uma referência temporal (`eventNPT`) e um campo de dados privados (`privateDataBytes`). Com isso, é possível encapsular as mensagens LCM e LDM no campo de dados privados (`privateDataBytes`) e as

informações de sincronização no campo de referência temporal (*eventNPT*), e embarcar, posteriormente, essas informações num fluxo MPEG-2 TS.

Tabela 5.4: Sintaxe dos eventos DSM-CC

DSM-CC stream events{	
descriptor_tag	8 bits
descriptor_length	8 bits
eventID	8 bits
reserved	31 bits
eventNPT	33 bits
privateDataBytes	8-1952 bits
}	

Na próxima seção, serão apresentados os detalhes de implementação dos componentes do LibrasTV para o SBTVD.

5.1.2 Implementação dos Componentes

Conforme a Figura 5.2, no LibrasTV, alguns componentes da solução proposta são integrados na estação de TV (módulo "Tradutor de LIBRAS"), e outros são executados como uma aplicação interativa no receptor de TV Digital.

Integrados na estação de TV, os componentes de Filtragem, Extração de Legendas e Tradução Automática compõem o módulo "Tradutor de LIBRAS" e foram implementados utilizando a linguagem de programação C++ e considerando as características apresentadas nas Seções 4.1 e 4.2.

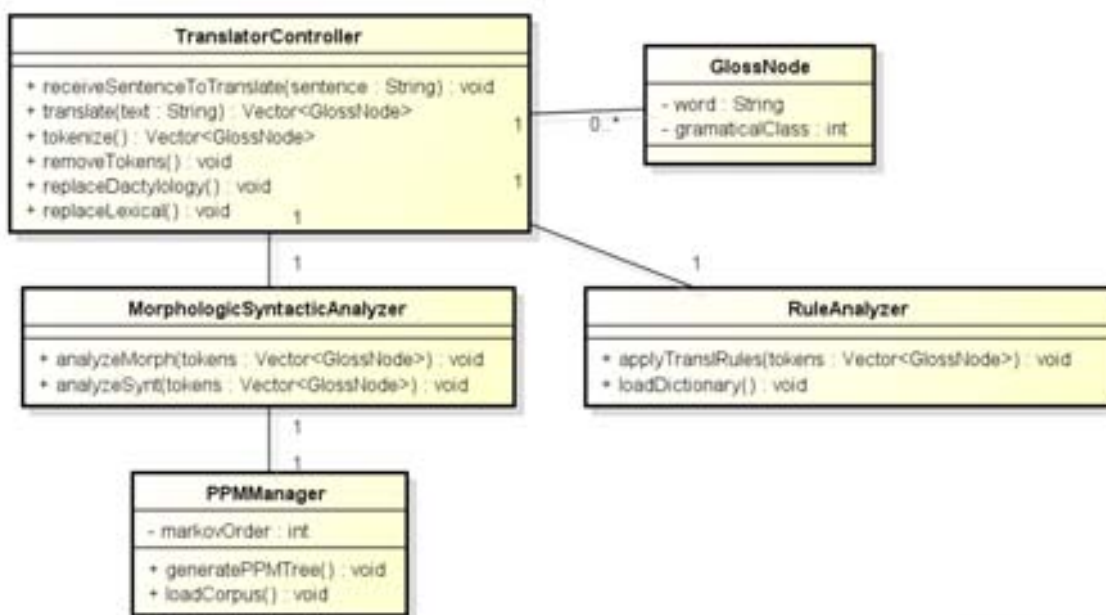
Os componentes de Filtragem e Extração de Legendas foram desenvolvidos com base nas definições das especificações MPEG-2 Sistemas [ISO/IEC 1996] e ABNT NBR 15606-1 [ABNT 2007], respectivamente. A especificação MPEG-2 Sistemas define como extrair os diversos fluxos elementares (áudio, vídeo, legendas, *closed caption*, etc.) de um fluxo MPEG-2 TS, enquanto que a ABNT NBR 15606-1 define, dentre outras coisas, como extrair textos a partir de um fluxo elementar de legenda ou *closed caption* no SBTVD. Esses componentes recebem um fluxo MPEG-2 TS, identificam os pacotes de legendas (Filtragem) e extraem as sentenças em Português e as informações de sincronização (etiquetas de tempo - *timestamps*) desses pacotes (Extração de Legendas).

O componente de Tradução Automática recebe as sentenças em Português e as traduz para uma seqüência de glosas em LIBRAS. Ele foi desenvolvido de acordo com o

diagrama de classes ilustrado na Figura 5.3.

A classe principal desse componente é a classe `TranslatorController`. Ela possui um método `receiveSentencesToTranslate()` que recebe sentenças em Português e utiliza os métodos `translate()`, `tokenize()`, `removeTokens()`, `replaceDactylogy()` e `replaceLexical()` para traduzir essa sentença para uma seqüência de glosas em LIBRAS. Ele também possui instâncias das classes `MorphologicSyntacticAnalyzer` e `RuleAnalyzer` que são utilizadas, respectivamente, para classificar os *tokens* morfo-sintaticamente e aplicar as regras de tradução armazenadas na base de tradução.

Figura 5.3: Diagrama de classe do componente de Tradução Automática



A classificação Morfológica e Sintática é realizada com base no corpus "Bosque"³ [Freitas et al. 2008]. Esse corpus foi desenvolvido pelo projeto "Floresta Sintática" [Freitas et al. 2008] e possui 9.368 sentenças e 186.000 palavras em língua portuguesa. Essas sentenças foram obtidas do jornal brasileiro "Folha de São Paulo" e do jornal português "Público", e foram classificadas morfológica e sintaticamente por lingüistas. Na implementação deste classificador, no entanto, apenas as sentenças do corpus em Português Brasileiro foram utilizadas nessa implementação.

A classe `MorphologicSyntacticAnalyzer` possui uma instância da classe `PPMManager`. Essa classe, inicialmente, utiliza o algoritmo do PPM-C [Moffat 1990] para construir um

³www.linguateca.pt/floresta/corpus.html#bosque

modelo estatístico da língua portuguesa baseado nas sentenças do Bosque. Em seguida, o PPM-C [Moffat 1990] é aplicado para classificar os *tokens* morfológica e sintaticamente. O modelo do PPM foi implementado utilizando uma cadeia de Markov de ordem 5. Esse valor de ordem foi definido empiricamente com o objetivo de manter um bom equilíbrio entre precisão e tempo de execução.

A classe RuleAnalyzer aplica as regras de tradução definidas na base de "Regras de Tradução" (ver Figura 4.2) para a seqüência de *tokens* e utiliza um dicionário de Português-LIBRAS para realizar a substituição léxica. Conforme mencionado na Seção 4.2., na substituição léxica, alguns *tokens* são substituídos para adaptar o significado da sentença rescrita para LIBRAS, uma vez o vocabulário de LIBRAS é menor do que o vocabulário da língua portuguesa [Santos et al. 2009]. Esse dicionário Português-LIBRAS foi desenvolvido em duas partes. A primeira parte foi extraída do "Dicionário Ilustrado de Libras do Estado de São Paulo" que possui 43.606 entradas, 3.340 imagens e 3.585 vídeos, onde um intérprete representa os sinais de LIBRAS. A segunda parte foi gerada por especialistas em LIBRAS, considerando a variação de inflexão verbal entre as línguas, isto é, que os verbos conjugados em Português são traduzidos para sua forma infinitiva em glosa. O dicionário Português-LIBRAS completo utilizado possui 295.451 entradas.

Para codificar a seqüência de glosas no fluxo MPEG-2 TS, um componente de Codificação também foi implementado no "Tradutor de LIBRAS". Esse componente também foi desenvolvido utilizando a linguagem C++ e funciona da seguinte forma. Inicialmente, ele recebe a seqüência de glosas do componente de Tradução Automática e gera as mensagens LCM e LDM de acordo com o protocolo definido na Seção 5.1.1. Essas mensagens são então encapsuladas em eventos DSM-CC juntamente com as informações de sincronização (etiquetas de tempo - *timestamps*) e são, em seguida, empacotadas em pacotes TS para multiplexação.

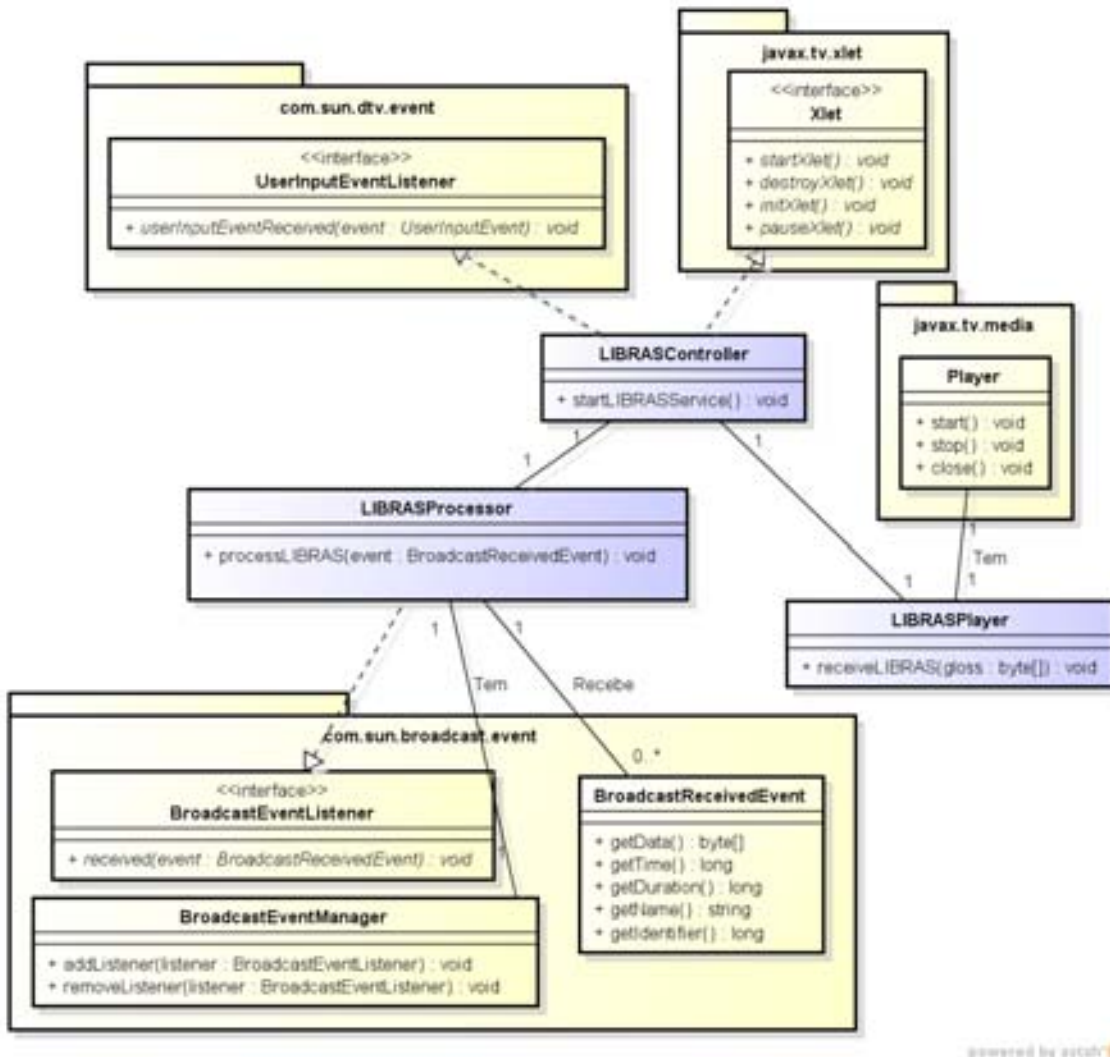
O multiplexador então recebe os pacotes TS gerados pelo componente de Codificação, multiplexa esses pacotes no fluxo MPEG-2 TS juntamente com os pacotes de áudio, vídeo e dados, e encaminha o fluxo MPEG-2 TS resultante para ser modulado e transmitido pela rede de difusão.

No lado do receptor, uma aplicação Ginga-J⁴, agrupa as funcionalidades dos componentes de Animação e Distribuição para gerar e exibir de forma sincronizada, respectivamente, o vídeo (ou trilha) de LIBRAS. Essa aplicação também possui um módulo de Decodificação, responsável por decodificar os eventos DSM-CC e extrair a seqüência de glosas e informações de sincronização encapsuladas nesses eventos. Na Figura 5.4, é

⁴Ginga-J é parte procedural do middleware Ginga, o middleware do SBTVD. As APIs do Ginga-J são baseadas na linguagem de programação Java [Souza Filho et al. 2007]

ilustrado o diagrama de classes dessa aplicação.

Figura 5.4: Diagrama de classes da aplicação LibrasTV.



A classe principal dessa aplicação é a classe `LIBRASController`. Essa classe possui instâncias das classes `LIBRASProcessor` e `LIBRASPlayer`, que implementam as funcionalidades dos componentes de Decodificação e de Animação e Distribuição, respectivamente. Ela também possui instâncias de outras classes Ginga, como, por exemplo, as classes `javax.tv.xlet.Xlet` e `com.sun.dtv.ui.event.UserInputEventListener` que são responsáveis por controlar as entradas do usuário pelo controle remoto e por gerenciar o ciclo de vida da aplicação, respectivamente.

A classe `LIBRASProcessor` implementa a interface `com.sun.dtv.event.BroadcastEventListener`

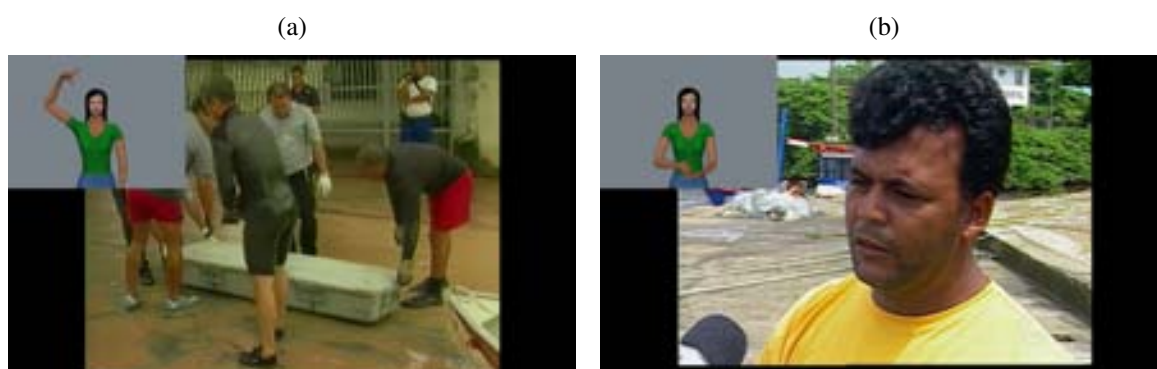
e utiliza instâncias da classe `com.sun.dtv.broadcast.event.BroadcastEventManager` para se registrar e capturar os eventos DSM-CC recebidos pela aplicação. Dessa forma, a `LIBRASProcessor` pode manipular objetos do tipo `com.sun.dtv.broadcast.BroadcastReceivedEvent`, que são instâncias dos eventos DSM-CC nas aplicações Ginga-J.

A classe `LIBRASPLayer` então recebe a seqüência de glosas e as informações de sincronização, extraídas pelas instâncias da classe `LIBRASProcessor`, gera um vídeo de LIBRAS (Animação) com o auxílio do Dicionário de LIBRAS e exibe essa trilha como uma janela de LIBRAS de forma sincronizada sobre o vídeo do programa (Distribuição). Para isso, essa classe utiliza a API "*Java Media Framework (JMF) 1.0*", disponível no pacote `javax.media` do Ginga-J.

Nessa versão do protótipo, o Dicionário de LIBRAS foi armazenado em um dispositivo de memória externa (USB - *Universal Serial Bus*) do receptor e os sinais do Dicionário foram gerados com base no modelo do avatar-3D apresentado na Seção 4.5.4. Para suavizar a transição entre sinais consecutivos, uma configuração (posição) neutra foi utilizada no início e no final da representação de todos os sinais do Dicionário de LIBRAS. Essa posição foi definida de acordo com a sugestão de intérpretes de LIBRAS e posiciona as mãos e os braços de forma estendida em linha reta para baixo e com uma expressão facial neutra (isto é, sem aplicar qualquer movimento nos ossos da face do avatar-3D).

A Figura 5.5, ilustra algumas capturas de tela da janela de LIBRAS gerada pelo protótipo apresentado. Essa aplicação foi testada e validada numa versão adaptada do *Openginga*⁵, implementação de código aberto do *middleware* Ginga.

Figura 5.5: Capturas de tela da execução da aplicação LibrasTV sobre o Openginga



Na próxima seção será apresentado o protótipo da solução proposta para Web, denominado *LibrasWeb*.

⁵O *Openginga* é uma implementação de código aberto do *middleware* Ginga disponível em gingacdn.lavid.ufpb.br/projects/openginga

5.2 LibrasWeb

O protótipo da solução proposta desenvolvido para Web, denominado LibrasWeb, foi implementado com todos os seus componentes sendo executados em um (ou mais) servidor(es) na nuvem. Nesse contexto, o LibrasWeb basicamente recebe um conteúdo multimídia de entrada, gera automaticamente um vídeo (trilha) de LIBRAS a partir das legendas desse conteúdo e mixa o vídeo de LIBRAS nesse conteúdo para torná-lo acessível.

Uma característica importante dessa abordagem, é que ela permite enxergar a solução proposta como um "serviço na nuvem" que torna os conteúdos multimídia acessíveis para surdos ("Acessibilidade como um Serviço - *Accessibility as a Service - AaaS*") [Araújo et al. 2012]. Na Figura 5.6, é ilustrado a arquitetura do LibrasWeb explorando esse conceito.

Figura 5.6: Arquitetura geral de integração do LibrasWeb na nuvem



De acordo com a Figura 5.6, o usuário acessa o LibrasWeb através de uma interface Web submetendo um conteúdo multimídia (vídeo) de entrada. Ao receber o conteúdo, o LibrasWeb gera automaticamente um vídeo (trilha) de LIBRAS e mixa esse vídeo no conteúdo multimídia de entrada, tornando-o acessível.

Nas próximas subseções será descrita a implementação dos componentes do LibrasWeb e a estratégia utilizada para integrá-lo na nuvem.

5.2.1 Implementação dos Componentes

Conforme comentado anteriormente, no LibrasWeb, os componentes da solução proposta são executados em um (ou mais) servidor(es), provendo uma espécie de serviço de geração de conteúdos acessíveis na Web. Mais especificamente, o LibrasWeb recebe um fluxo multimídia de entrada, gera uma trilha de LIBRAS que é mixada no conteúdo

multimídia e devolvida para os clientes. Todos os componentes desse protótipo foram desenvolvidos na linguagem de programação C++ e considerando as características apresentadas no Capítulo 4.

Para os componentes de Filtragem, Extração de Legendas e Tradução Automática, a mesma implementação dos componentes desenvolvida no LibrasTV (ver Seção 5.1.2), foi utilizada no LibrasWeb. Esses componentes recebem um arquivo multimídia em formato MPEG-2 TS, identificam os pacotes de legendas (Filtragem) ou utilizam um arquivo de legenda separado (em formato Vobsub⁶), extraem as sentenças em Português e as informações de sincronização a partir do fluxo ou do arquivo de legenda (Extração de Legenda) e traduzem essas sentenças para uma seqüência de glosas em LIBRAS (Tradução Automática).

O componente de Animação recebe a seqüência de glosas do componente de Tradução Automática e procura a representação visual (vídeos) desses sinais no Dicionário de LIBRAS para gerar o vídeo de LIBRAS. Os sinais do Dicionário de LIBRAS foram gerados a partir do modelo do avatar-3D descrito na Seção 4.5.4. Além de aplicar a configuração neutra no início e no final da representação de cada sinal do Dicionário de LIBRAS, um "sinal neutro" com o avatar-3D na posição neutra também foi desenvolvido para o LibrasWeb. Esse sinal possui a duração de 1 (um) segundo e é exibido repetidamente pelo componente de Animação durante os intervalos de silêncio e entre o início do vídeo e a exibição da primeira legenda.

Para sincronizar o vídeo de LIBRAS gerado com o conteúdo multimídia de entrada, o componente de Animação extrai o primeiro relógio global de referência do fluxo de entrada, isto é, o primeiro PCR (*Program Clock Reference*) do fluxo MPEG-2 TS. Esse relógio de referência é então utilizado como o primeiro relógio do vídeo de LIBRAS. As etiquetas de tempo (*timestamps*) de todos os sinais de LIBRAS são então geradas com base nesse PCR e na etiqueta de tempo (*Presentation Timestamps - PTS*) do pacote de legenda relacionado. No final do processamento, o componente de Animação gera um fluxo de vídeo de LIBRAS completo.

O componente de Distribuição recebe esse vídeo de LIBRAS, juntamente com o conteúdo multimídia de entrada, e sobrepõe o vídeo de LIBRAS sobre o conteúdo multimídia de entrada, mixando os dois conteúdos. Para sobrepor o vídeo de LIBRAS de forma síncrona, o primeiro passo aplicado pelo componente de Distribuição é igualar a taxa de quadros (número de quadros por segundo) dos dois vídeos. Após essa tarefa, o vídeo de LIBRAS é sobreposto (mixado) ao vídeo de entrada, com base em um conjunto de

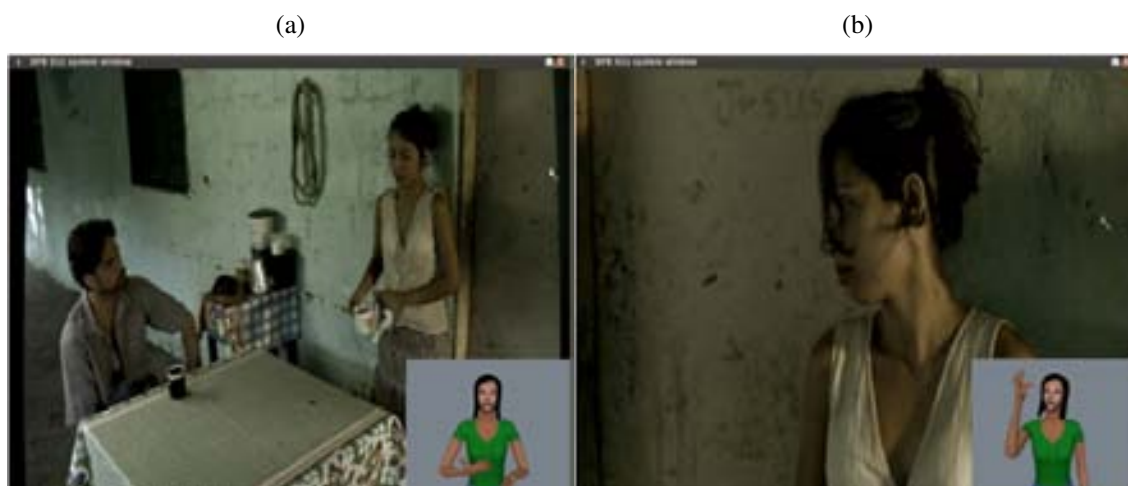
⁶O Vobsub é um formato de legendas frequentemente utilizado na Web, que possibilita o acréscimo, fácil, de legendas a um arquivo

parâmetros de tamanho e posição da janela de sobreposição fornecidos pelo usuário.

O processo de ajuste da taxa de quadros e mixagem foi implementado utilizando o FFmpeg⁷, uma ferramenta de código aberto desenvolvida para gravar, manipular, converter e transmitir fluxos de áudio e vídeo. Como resultado desse processo, o componente gera um novo arquivo de vídeo, onde o vídeo de LIBRAS é sobreposto ao conteúdo multimídia original (de entrada), tornando-o acessível.

A Figura 5.7 ilustra uma captura de tela dos vídeos gerados pelo LibrasWeb.

Figura 5.7: Capturas de tela do vídeo gerado pelo LibrasWeb



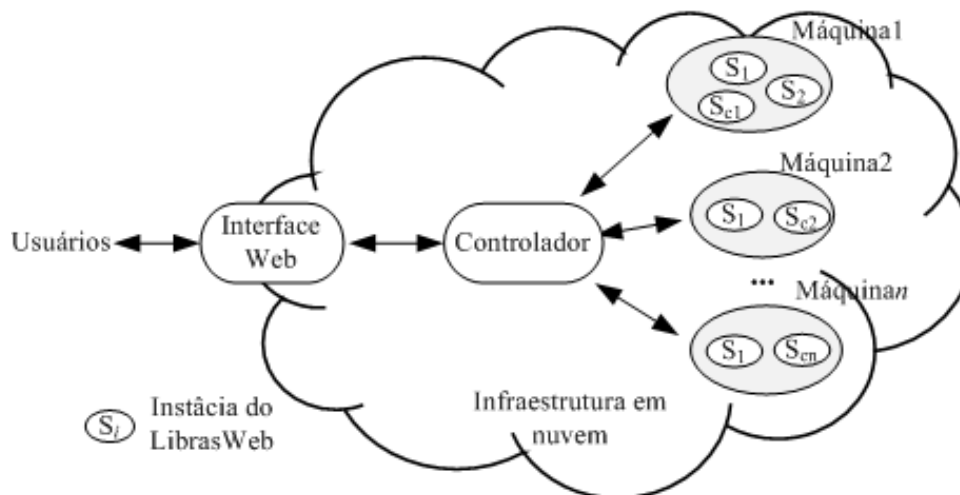
5.2.2 Integração na Nuvem

Uma alternativa interessante para integrar o LibrasWeb na nuvem, é embarcar a solução num provedor de computação em nuvem (*cloud computing*). Essa integração permite explorar as principais características do modelo da computação em nuvem, que incluem a alta escalabilidade, o provisionamento dinâmico de recursos e a facilidade de acesso [Zhang et al. 2010], dentre outras. A Figura 5.8 ilustra o modelo utilizado para embarcar o LibrasWeb em uma plataforma de computação em nuvem (provedor de infraestrutura - *Infrastructure as a Service (IaaS) provider*).

De acordo com a Figura 5.8, os usuários acessam o serviço através de uma interface Web e submetem seus conteúdos multimídia para processamento. Um componente Controlador então recebe essa requisição e escalona as requisições do usuário em um conjunto de máquinas de processamento. Ele então verifica a carga de processamento do sistema.

⁷ffmpeg.org

Figura 5.8: Visão esquemática da integração do LibrasWeb em uma infraestrutura de computação em nuvem.



Se a carga estiver abaixo de um determinado limite, ele escalona a execução da requisição para uma das máquinas virtuais providas pelo IaaS *provider*. Caso contrário, o Controlador requisita uma nova máquina virtual e realiza o escalonamento conforme descrito anteriormente. Em resumo, a estratégia de distribuição da carga de processamento é baseada no algoritmo de escalonamento *round-robin*, onde a requisição é atribuída a máquina que possui o menor número de instâncias em execução.

Finalmente, sempre que o processamento de uma requisição é concluído, o Controlador verifica se a carga do sistema está abaixo de um determinado limiar pré-definido. Caso esteja, as máquinas ociosas são encerradas até que a carga de processamento esteja acima desse limiar.

A interface Web desse serviço foi implementada utilizando a linguagem de programação Ruby e a tecnologia Adobe Flash Player⁸. Ela explora o conceito de vídeo interativo, onde as interações são baseadas em vídeos interativos produzidos com intérpretes de LIBRAS. A partir da interação direcionada pelos vídeos interativos, o usuário submete o conteúdo multimídia, juntamente com parâmetros como tamanho e posição da janela, gerando uma requisição do serviço que será tratada pelo Controlador.

O Controlador foi implementado na linguagem de programação C++ e gerencia um conjunto de n instâncias de máquinas virtuais no Amazon *Elastic Compute Cloud* (Amazon EC2)⁹, um provedor de infraestrutura (IaaS *provider*). O Amazon EC2 possui um conjunto de máquinas virtuais de propósito geral e permite que usuários na nuvem tenham

⁸www.adobe.com/br/products/flashplayer.html

⁹aws.amazon.com/ec2/

total controle sobre instâncias de servidores em *data centers* remotos.

Para executar o escalonamento proposto, o Controlador utiliza um vetor com n elementos, onde cada elemento representa um contador do número de atividades em execução em cada instância alocada no Amazon EC2. Sempre que o usuário faz uma nova requisição, o Controlador cria uma nova instância do LibrasWeb (isto é, uma nova tarefa), que será executada na instância do Amazon EC2 com o menor número de tarefas em execução. É importante ressaltar, no entanto, que a avaliação dos aspectos relacionados à estratégia de distribuição, como, por exemplo, a escalabilidade e o provisionamento de recursos estão fora do escopo deste trabalho, sendo deixado para trabalhos futuros.

A Figura 5.9 ilustra duas capturas de tela do protótipo de integração do LibrasWeb na nuvem. Após a configuração dos parâmetros na interface Web do serviço, uma instância do LibrasWeb é criada e executada, produzindo uma cópia acessível do conteúdo, que é devolvida ao usuário. Na Figura 5.9a, é ilustrada a configuração do parâmetro posição da janela de LIBRAS no vídeo interativo com intérprete de LIBRAS. De acordo com a Figura 5.9a, quatro opções de posição da janela de LIBRAS (canto superior esquerdo, canto superior direito, canto inferior esquerdo e canto inferior direito) são apresentadas ao usuário sobre o vídeo e o usuário seleciona a opção desejada clicando sobre uma das quatro posições oferecidas. Na Figura 5.9b, é ilustrada a cópia do conteúdo acessível gerado pelo protótipo e apresentada na tela ao usuário. O usuário também pode fazer, opcionalmente, o *download* da nova cópia do conteúdo.

Figura 5.9: Capturas de tela do protótipo do LibrasWeb: (a) tela de seleção da posição da janela de LIBRAS; (b) tela de apresentação da cópia acessível do conteúdo.



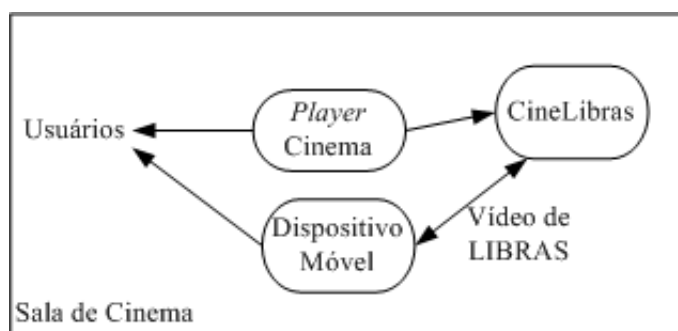
Na próxima subseção será apresentado o protótipo da solução proposta para Cinema Digital, denominado CineLibras.

5.3 CineLibras

O CineLIBRAS, protótipo da solução proposta desenvolvido para Cinema Digital, foi implementado considerando a geração automática das trilhas de LIBRAS na própria sala de Cinema. Para isso, o protótipo é executado em um servidor integrado ao dispositivo de apresentação do vídeo (*Player*) na sala de Cinema.

A proposta é que o vídeo (trilha) de LIBRAS seja gerado automaticamente a partir das legendas do filme na sala de Cinema, e seja transmitido para os dispositivos móveis (*tablets* ou *smartphones*, por exemplo) dos usuários surdos de forma independente, permitindo que eles acompanhem a tradução para LIBRAS nos seus próprios dispositivos dentro de um ambiente compartilhado de exibição. Essa adaptação pode ser feita, inclusive, adicionando assentos especiais para surdos nas salas de Cinema Digital, onde os dispositivos móveis estariam embutidos nos assentos e programados para receber a tradução para LIBRAS. A Figura 5.10 ilustra a arquitetura de implementação desse protótipo.

Figura 5.10: Arquitetura geral de integração do CineLibras



Conforme pode ser observado na Figura 5.10, o protótipo integrado com o *Player* de Cinema Digital recebe o fluxo de legenda em Português do filme. Esses fluxos são extraídos pelo protótipo, que gera um vídeo (trilha) de LIBRAS a partir do texto extraído da legenda e encaminha (distribui) esse vídeo para os dispositivos móveis dos usuários.

Os componentes do CineLIBRAS foram implementados de forma similar à implementação do LibrasWeb (apresentados na Seção 5.2.1), com mudanças apenas nos componentes de Filtragem, Extração de Legendas e Distribuição. Os componentes de Filtragem e Extração de Legendas extraem as legendas do filme no formato DCP (*Digital Cinema Package*)¹⁰ [Digital Cinema Initiatives 2008], formato utilizado para codificar legendas em Cinema Digital, e o componente de Distribuição transmite o vídeo de LIBRAS,

¹⁰O DCP é uma coleção de arquivos digitais utilizadas para armazenar e transmitir fluxos de áudio, vídeo e dados, incluindo legendas, em Cinema Digital

gerado em tempo real pelo componente de Animação, via HTTP (*streaming* HTTP - *HyperText Transfer Protocol*) para os dispositivos móveis dos usuários conectados. Os problemas recorrentes de atraso e variação de retardo de transmissão nesse tipo de solução não tendem a ser um problema, uma vez que, conforme mencionado anteriormente, de acordo com a norma ABNT NBR 15290 [ABNT 2005], são tolerados quatro segundos de atraso em sistemas de *closed caption* ao vivo.

No dispositivo móvel, os usuários se conectam ao CineLIBRAS utilizando *Players* com suporte a recepção de fluxos MPEG-2 TS via *streaming* HTTP. Alguns testes preliminares foram realizados em dispositivos móveis com Sistema Operacional Android 2.2, 2.3 e 3.0 e utilizando o VPlayer¹¹, um *player* de vídeo disponível para essas plataformas com suporte a recepção de vídeos via *streaming* HTTP.

Na Figura 5.11, é ilustrada uma demonstração do CineLIBRAS realizada no XIII Workshop da RNP (WRNP), evento realizado na cidade de Ouro Preto-MG, dentro do XXX Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos (SBRC 2012)¹².

Figura 5.11: Demonstração do CineLIBRAS no XIII WRNP. O vídeo do Cinema aparece ao fundo, enquanto que o vídeo de LIBRAS é gerado pelo protótipo e transmitido para o dispositivo móvel do usuário.



¹¹vplayer.net

¹²sbr2012.dcc.ufmg.br

5.4 WikiLIBRAS

O WikiLIBRAS foi implementado conforme descrito na Seção 4.5.1. A interface Web do WikiLIBRAS foi desenvolvida utilizando a linguagem de programação PHP com o auxílio da tecnologia Adobe Flash Player[Adobe 2012a]. Uma vez que os surdos têm, em geral, dificuldades para ler e escrever em Português [Stumpf 2000], essa interface foi projetada para ser acessível e inteligível para o seu público alvo (surdos brasileiros e intérpretes de LIBRAS), tendo como foco a utilização de elementos gráficos e animações, e explorando o mínimo possível de elementos textuais.

Quando o usuário seleciona a opção de "Cadastrar um Sinal", uma janela é apresentada para o usuário configurar o nome do sinal (glosa), o tipo de movimento e, o número de repetições daquele movimento (ver Figura 5.12). Após a seleção de cada parâmetro, as opções selecionadas pelos usuários são apresentadas no lado direito da tela ("*Preview*"), gerando um *feedback* para o usuário sobre as opções selecionadas. Adicionalmente, uma animação é gerada ilustrando o tipo de trajetória do movimento selecionado. Por exemplo, se o usuário seleciona um movimento do tipo retilíneo, uma animação seria apresentada com a mão direita realizando um movimento retilíneo, o que torna a configuração desse parâmetro mais inteligível para os usuários.

Figura 5.12: Captura da tela principal de cadastro de um sinal no WikiLIBRAS.



Em seguida, o usuário é direcionado para configurar os fonemas de configuração de mão, orientação mão, ponto de articulação e expressão facial. Na Figura 5.13, são apresentadas algumas capturas de tela dessas etapas.

Conforme pode ser observado na Figura 5.13, nestas telas, um conjunto de imagens

Figura 5.13: Capturas de tela para configuração dos fonemas: (a) expressão facial, (b) configuração de mão, (c) orientação da palma da mão e (d) ponto de articulação.



(opções) associadas com os fonemas são apresentados aos usuários. O usuário deve então selecionar a opção que está relacionada com o sinal que está sendo cadastrado. Adicionalmente, uma *timeline* é apresentada na parte de baixo da tela com as opções já selecionadas pelos usuários. Caso o usuário deseje modificar algum dos parâmetros já configurados, ele seleciona o parâmetro na *timeline* e atualiza o seu valor.

Finalmente, após selecionar todos os parâmetros, eles são encaminhados para o Gerador de Descrição de Sinais, que os converte numa representação XML de acordo com a LDS.

Quando o usuário seleciona a opção de "Cadastrar" ou "Editar" uma Regra, ele é encaminhado para uma tela de cadastro ou consulta das regras de tradução, podendo configurar os parâmetros relacionados a uma regra de tradução (ver Figura 5.14). As regras implementadas no componente de Tradução Automática dos protótipos LibrasTV e LibrasWeb foram definidas por intérpretes de LIBRAS utilizando essas opções dos sistemas. De forma similar ao cadastro de novos sinais, após a configurações dos parâmetros da regra, eles são encaminhados para o módulo Gerador de Descrição de Regras que os converte

numa representação XML de acordo com a LDRT.

Figura 5.14: Capturas de tela para (a) Cadastro e (b) Edição de Regras.



Os módulos Gerador de Descrição de Sinais e Gerador de Descrição de Regras, responsáveis por gerar, respectivamente, a representação XML das regras de tradução e sinais, também foram desenvolvidos utilizando a linguagem PHP. Quando o usuário cadastra ou edita uma regra de tradução, a representação XML dessa regra é salva e enviada para a etapa de supervisão. Quando o usuário cadastra um sinal, essa representação é encaminhada para o servidor que converte essa representação para vídeo através dos módulos *Parser* e *Renderizador*.

Os módulos *Parser* e *Renderizador*, responsáveis, respectivamente, por converter a representação XML em uma representação intermediária e gerar um vídeo para o sinal a partir dessa representação intermediária e com base no modelo do avatar-3D (descrito na Seção 4.5.4), foram implementados através de *scripts* na linguagem de programação Python. Esses *scripts* são responsáveis por interpretar a linguagem intermediária gerada pelo *Parser*, configurar os fonemas e renderizar os sinais utilizando as bibliotecas de poses pré-gravadas (mencionadas na Seção 4.5.3). Para os fonemas configuração de mão, orientação da mão, ponto de articulação e expressão facial, bibliotecas de poses foram criadas com as coordenadas de rotação e localização dos ossos utilizados em cada uma das poses. Essas bibliotecas auxiliam na construção dos *scripts* de renderização e facilitam a edição dos parâmetros de configuração do sinal definidos na LDS. Para cada expressão facial, por exemplo, é necessário configurar as rotações e localizações dos 21 ossos localizados no rosto do avatar-3D.

Após gerar o vídeo do sinal, o sistema redireciona o usuário para uma tela que apresenta o vídeo do sinal para o usuário (ver Figura 5.15). Após a confirmação do usuário,

esse vídeo é enviada para a etapa de supervisão, encerrando o processo de interação do usuário.

Figura 5.15: Captura de tela do vídeo do sinal gerado pelo WikiLIBRAS.



Na próxima seção serão apresentados alguns testes preliminares conduzidos com os protótipos LibrasTV e LibrasWeb. Nesses testes preliminares alguns aspectos da solução proposta são previamente avaliados. Uma avaliação mais profunda e planejada da solução, no entanto, será descrita e desenvolvida no Capítulo 6.

5.5 Testes Preliminares com os Protótipos

5.5.1 LibrasTV

Após implementar o protótipo do LibrasTV, alguns testes computacionais objetivos foram realizados com o objetivo de avaliar de forma preliminar alguns aspectos da solução proposta, como, por exemplo, a qualidade de tradução, o atraso na tradução e a largura de banda do canal utilizada pelo fluxo de LIBRAS codificado. O ambiente utilizado no teste e os principais resultados obtidos serão discutidos nas próximas subseções.

Ambiente do Teste

Para realizar os primeiros testes com o protótipo, dois computadores (mini-PCs) com processador Intel Dual Core T3200 2GHz e 4GB de memória RAM, juntamente com alguns equipamentos de estação de TV profissionais, como, por exemplo, um *Streamer* de Vídeo, um Gerador de Carrossel (*Data Stream*), e um Multiplexador foram utilizados.

O primeiro computador foi utilizado para executar o módulo "Tradutor de LIBRAS" do protótipo e o segundo computador foi utilizado para executar a aplicação interativa LibrasTV (conforme descrito na Seção 5.1.2). Uma versão do OpenGinga foi instalada no segundo computador para tornar possível a execução da aplicação. O sistema operacional utilizado nos dois computadores foi o Linux Ubuntu 10.0.4, kernel 2.6.32.

Esses computadores e equipamentos de TV Digital foram integrados e organizados da seguinte forma:

- O *Streamer* de Vídeo transmite via *streaming* UDP alguns vídeos MPEG-2 TS de teste com fluxo de legenda embutido para o Tradutor de LIBRAS e para o Multiplexador;
- O "Tradutor de LIBRAS" recebe o fluxo MPEG-2 TS do *Streamer* de Vídeo e gera um fluxo de LIBRAS codificado, isto é, a seqüência de glosas e as informações de sincronização codificadas nas mensagens LDM e LCM e encapsuladas nos eventos DSM-CC;
- O Gerador de Carrossel gera um fluxo de dados com a aplicação interativa do LibrasTV;
- O Multiplexador recebe o fluxo MPEG-2 TS, o fluxo de dados com a aplicação do LibrasTV e o fluxo de LIBRAS codificado, (re)multiplexa-os em um único fluxo MPEG-2 TS e transmite esse fluxo via *streaming* UDP para a máquina do OpenGinga (segundo computador);
- O OpenGinga recebe o fluxo MPEG-2 TS via *streaming* UDP, extrai e executa o vídeo do programa e a aplicação do LibrasTV que gera a trilha (janela) de LIBRAS a partir do fluxo de LIBRAS codificado.

Os fluxos de vídeo MPEG-2 TS com legendas utilizados nesses testes preliminares são apresentados na Tabela 5.5.

Tabela 5.5: Fluxos de Vídeo MPEG-2 TS com legendas utilizados nos testes

Vídeo	Duração	Descrição
Vídeo1	26 seg	Esse vídeo é parte de um programa de notícias veiculado em 14 de Outubro de 2008 na TV Globo, uma rede de TV Brasileira
Vídeo2	79 seg.	Esse vídeo é parte de um programa de notícias veiculado em 31 de Janeiro de 2007 na TV Globo

Tradução Automática

Inicialmente, uma técnica de validação cruzada múltipla foi aplicada para avaliar o desempenho do classificador morfo-sintático embutido no componente de Tradução Automática. Na técnica de validação cruzada múltipla, o conjunto de dados é dividido em K partes iguais. O modelo é então construído (ou treinado) com todas as partes, exceto uma que é utilizada para calcular o erro de validação. O procedimento é repetido K vezes e, em cada iteração, uma parte diferente do conjunto de dados é usada para validação.

Nesse teste, a técnica de validação cruzada múltipla foi aplicada sobre o nosso conjunto de dados, o corpus "Bosque", dividindo-o em 10 partes iguais. O procedimento foi repetido 10 vezes e, em cada execução, 9 partes foram incluídas no conjunto de treinamento do classificador morfo-sintático e a parte remanescente (uma parte diferente em cada execução) foi utilizada para avaliar o desempenho do classificador. O percentual de classificações corretas obtidas para cada execução é ilustrado na Tabela 5.6.

Tabela 5.6: Medidas do percentual de classificação correta das classificações morfo-sintáticas sobre o corpus "Bosque"

Execução	Classificações corretas
1	82,81%
2	83,50%
3	82,85%
4	83,07%
5	81,90%
6	79,72%
7	81,15%
8	81,44%
9	81,34%
10	81,01%
Média	81,88%

Conforme pode ser observado na Tabela 5.6, o classificador obteve uma média de acerto de 81,88% na classificação dos conjuntos de validação. Isso implica que o erro médio de classificação foi inferior a 20%.

Em seguida, alguns testes objetivos foram conduzidos para avaliar a qualidade da saída do componente de Tradução Automática. Nesses testes, os valores de WER (*Word error rate*) e BLEU (*Bilingual Evaluation Understudy*) [Papineni et al. 2001] foram calculados. Essas medidas foram escolhidas porque elas também foram utilizadas em trabalhos ante-

riores [San-segundo et al. 2008b][Su & Wu 2009][Wu et al. 2007], embora em domínios diferentes (ver Capítulo 3).

Para realizar esses testes, inicialmente, dois intérpretes de LIBRAS traduziram todas as sentenças do corpus "Bosque" para uma seqüência de glosas em LIBRAS, gerando uma tradução de referência para todas as sentenças do corpus. Em seguida, as sentenças do corpus foram traduzidas usando o protótipo do LibrasTV e os valores de WER e BLEU foram calculados com base nas traduções de referência geradas pelos intérpretes. Os valores de WER e BLEU também foram calculados para uma solução baseada em tradução direta (isto é, que não faz as adaptações gramaticais para a gramática de LIBRAS). Com isso, é possível comparar esses resultados com os resultados de tradução da solução proposta. A Tabela 5.7 ilustra os valores percentuais de BLEU, considerando a precisão em 1-gramas (unigramas), 2-gramas (bigramas), 3-gramas (trigramas) e 4-gramas, e WER para as duas soluções.

Tabela 5.7: Valores de BLEU e WER para o LibrasTV e para uma solução baseada em Tradução Direta.

		LibrasTV	Tradução direta
BLEU	1-gramas	48,5%	40,7%
	2-gramas	30,1%	22,2%
	3-gramas	18,9%	11,4%
	4-gramas	12,0%	5,5%
WER		75,3%	87,7%

De acordo com a Tabela 5.7, nesses testes, as medidas do LibrasTV foram melhores do que as medidas da solução baseada em Tradução Direta para todos os tipos de n -gramas. Os valores de BLEU 4-gramas = 12% e WER = 75,3%, respectivamente, ajudam a avaliar como a tarefa de tradução sobre textos livres (domínio geral) é uma tarefa não-trivial. Deve ser ressaltado, no entanto, que o objetivo da solução proposta, não é substituir ou se equiparar a tradutores humanos. A idéia principal é reduzir as barreiras de acesso a informação, especialmente quando intérpretes não estão disponíveis. Além disso, esse resultado não é suficiente para concluir que a proposta de tradução é boa ou ruim. De acordo com Su & Wu (2009), a avaliação baseada em métricas objetivas não é suficiente para avaliar a qualidade de tradução para línguas de sinais, uma vez que essas línguas são gestuais e visuais. Em consequência disso, no Capítulo 6 serão apresentados alguns experimentos conduzidos com usuários para avaliar a qualidade de tradução.

Atraso e Largura de Banda

Alguns testes preliminares também foram realizados para estimar o tempo (atraso) de tradução do LibrasTV. Esses testes foram realizados com o Vídeo1 e Vídeo2 (ver Tabela 5.5). Para cada vídeo, o protótipo do LibrasTV foi executado em *loop* durante 4 minutos e o tempo de execução de cada componente do LibrasTV foi medido em milissegundos. O valor médio e o desvio padrão desses tempos são apresentados na Tabela 5.8.

Tabela 5.8: Medidas do tempo (atraso) médio de cada componente do LibrasTV

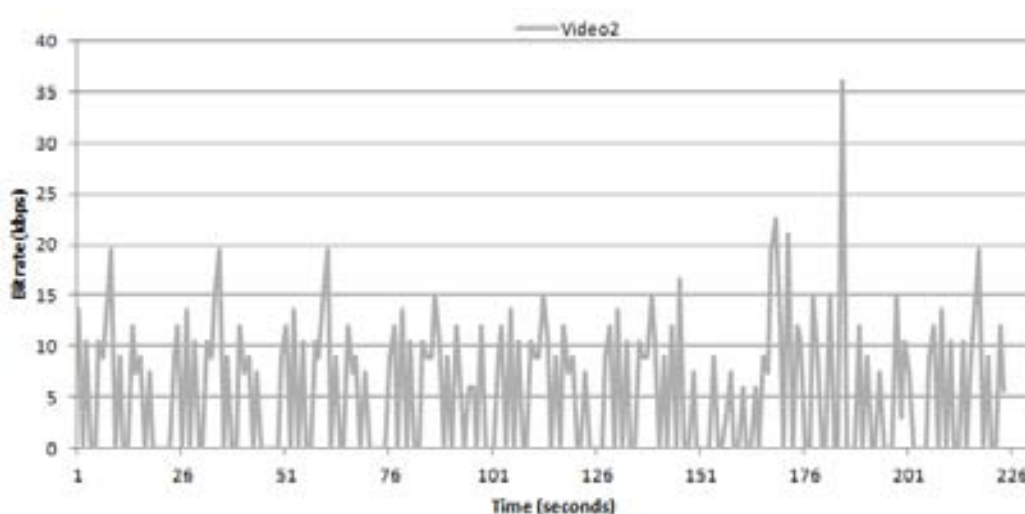
Vídeo	Componente	Média (ms)	Desvio Padrão (ms)
Vídeo1	Filtragem e Extração de Legendas	0,025	0,004
	Tradução Automática	0,881	0,334
	Codificação	0,798	0,113
	Decodificação	0,273	0,218
	Animação e Distribuição	40,214	12,130
	Total	42,191	-
Vídeo2	Filtragem e Extração de Legendas	0,023	0,003
	Tradução Automática	0,716	0,236
	Codificação	0,677	0,187
	Decodificação	0,255	0,243
	Animação e Distribuição	43,403	12,131
	Total	45,074	-

De acordo com a Tabela 5.8, o tempo médio necessário para executar todos os componentes (isto é, a soma dos tempos médios de Filtragem, Extração de Legendas, Tradução Automática, Codificação, Decodificação, Animação e Distribuição) foi inferior a 46 milissegundos. Considerando que o atraso máximo tolerado para sistemas de transmissão de *closed caption* é de 4 segundos [ABNT 2005], o baixo atraso médio obtido é um indício de que a solução proposta possui um bom desempenho de tradução e pode ser viável para traduções em tempo real. Para esse aspecto, um teste mais completo e envolvendo um cenário mais representativo será apresentado no Capítulo 6.

Por fim, um teste preliminar também foi realizado para avaliar a largura de banda utilizada pela solução. Nesse teste, o Vídeo1 e Vídeo2 foram transmitidos para o LibrasTV em seqüência por 2 iterações, e a taxa de bits (em Kbps) utilizada pelo fluxo de LIBRAS codificado (eventos DSM-CC com glosas encapsuladas) foi calculada. A taxa de bits (em Kbps) também foi calculada quando o vídeo com um intérprete de LIBRAS era transmitido para o usuário, em vez do fluxo de LIBRAS codificado. É importante ressaltar que

para os dois fluxos foi utilizado uma resolução padrão que corresponde a 1/4 da largura x 1/2 da altura do vídeo original. Essa resolução é a resolução mínima tolerável para janelas de LIBRAS, de acordo com a norma ABNT NBR 15290 [ABNT 2005]. As Figuras 5.16 e 5.17 mostram os resultados obtidos pelo LibrasTV e pela solução baseada em vídeos, respectivamente.

Figura 5.16: Largura de banda (em Kbps) utilizada pelo fluxo de LIBRAS codificado

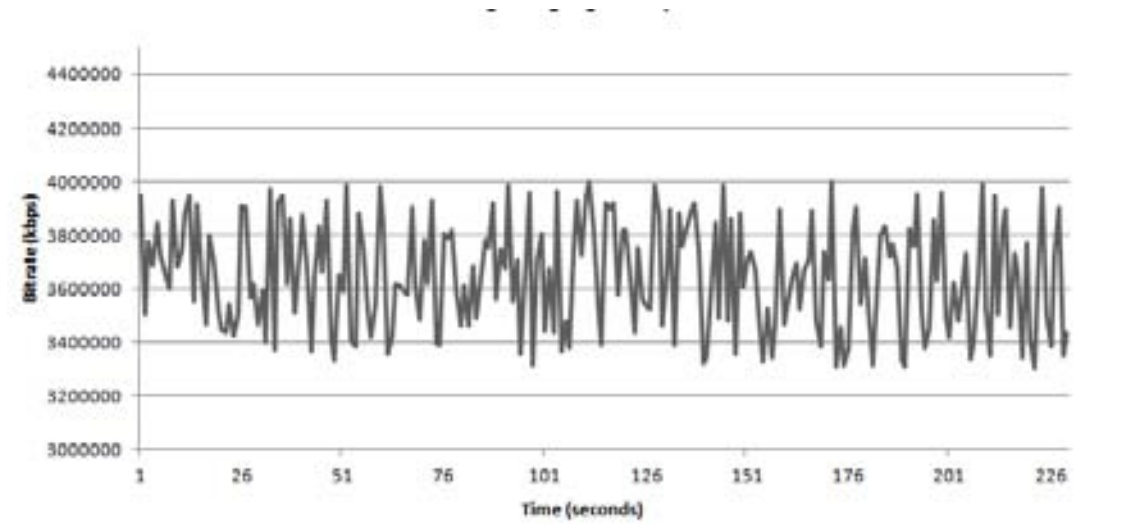


De acordo com as Figuras 5.16 e 5.17, pode-se observar a largura de banda de canal necessária para transmitir o fluxo de LIBRAS codificado (um valor médio de 5,64 Kbps e sempre inferior a 40 Kbps) foi bem inferior ao valor utilizado quando um fluxo de vídeo com intérprete de LIBRAS era utilizado (um valor médio de 3,64 Mbps). Um valor inferior de taxa de bits para o fluxo de LIBRAS codificado já era esperado, mas os testes indicaram que essa taxa é, em média, cerca de 600 vezes menor que o valor utilizado pelo fluxo de vídeo. O fluxo de vídeo com intérprete de LIBRAS utilizado no teste foi codificado utilizando um codificador de vídeo profissional MPEG-2. Outro fator importante a ser observado, é que a taxa do fluxo do vídeo foi alta mesmo durante os intervalos de silêncio.

5.5.2 LibrasWeb

Para avaliar a integração do LibrasWeb na nuvem, a solução foi embarcada no Amazon EC2, utilizando instâncias do tipo "EC2 Extra-large". Essas instâncias custam U\$

Figura 5.17: Largura de banda (em Kbps) utilizada pelo fluxo de vídeo com intérprete de LIBRAS



0,68 por hora¹³ e possuem 8 unidades de processamento EC2 (sendo 4 núcleos virtuais com 2 unidades de processamento EC2 por núcleo), 15 GB de memória RAM e 1.690 GB de espaço em disco.

Nesses testes preliminares, diferentes demandas de serviço foram simuladas nessas instâncias do Amazon EC2, e o tempo médio de resposta e o custo foram calculados em dois diferentes cenários. No primeiro cenário, o tempo médio de resposta e custo foram calculados utilizando apenas uma instância do EC2 e todas as requisições do usuário era direcionadas para essa instância (abordagem sequencial). No segundo cenário, o tempo médio de resposta e o custo foram calculados utilizando até 10 instâncias EC2 e as requisições foram distribuídas nessas instâncias de acordo com a estratégia descrita na Seção 5.2.2 (abordagem distribuída).

As simulações foram realizadas utilizando uma requisição uniforme, onde a entrada era um vídeo MPEG-2 TS em alta-definição com legendas em Português. O vídeo possuía uma taxa de quadro de 45 quadros por segundo e uma duração de 26 segundos. Cada simulação foi repetida três vezes para as duas abordagens e o tempo médio de resposta e o custo de cada abordagem foi medido. As Figuras 5.18 e 5.19 ilustram o tempo médio de resposta e o custo, respectivamente, para as requisições nos dois cenários descritos acima.

De acordo com as Figuras 5.18 e 5.19, a partir de 12 requisições, a abordagem distri-

¹³Os testes foram realizados em janeiro de 2012 e, na ocasião, eram esses os custos das instâncias (<http://aws.amazon.com/ec2/pricing>).

Figura 5.18: Tempo médio de resposta para diferentes requisições nas duas abordagens.

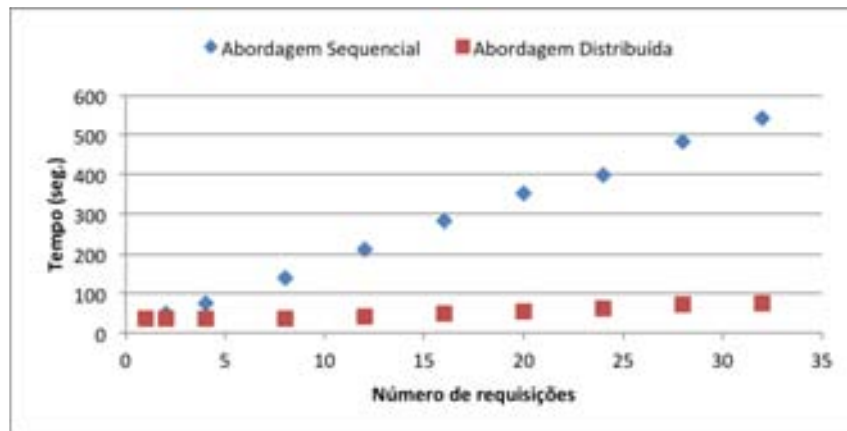
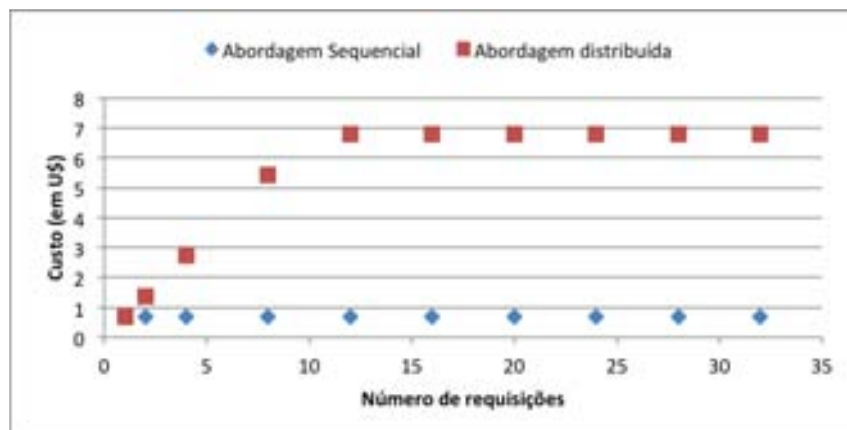


Figura 5.19: Custo médio para diferentes requisições nas duas abordagens.



buída obteve um desempenho cinco vezes melhor do que a abordagem sequencial. Considerando que, nesse exemplo, o custo cresceu apenas U\$ 6,12 por hora (isto é, U\$ 0,68 por hora * 9 instâncias = U\$ 6,12 por hora), a melhoria de desempenho da solução distribuída representou um ganho de desempenho de 81% por cada dólar extra investido. Para 24 requisições, o tempo médio de 397,13 segundos obtido pela abordagem sequencial foi reduzido para 60,92 segundos na abordagem distribuída, isto é, uma melhoria de desempenho (*speedup*) de 6,39 vezes. Nesse caso, o ganho de desempenho foi superior a 100% para cada dólar extra investido.

5.6 Considerações

Nesse capítulo foram apresentadas provas de conceito da solução proposta desenvolvidas para as plataformas de TV Digital, Web e Cinema Digital. As arquiteturas e adaptações da solução proposta para essas plataformas foram apresentadas e os detalhes de implementação dos protótipos foram descritos. Adicionalmente, alguns testes preliminares foram desenvolvidos para esses protótipos.

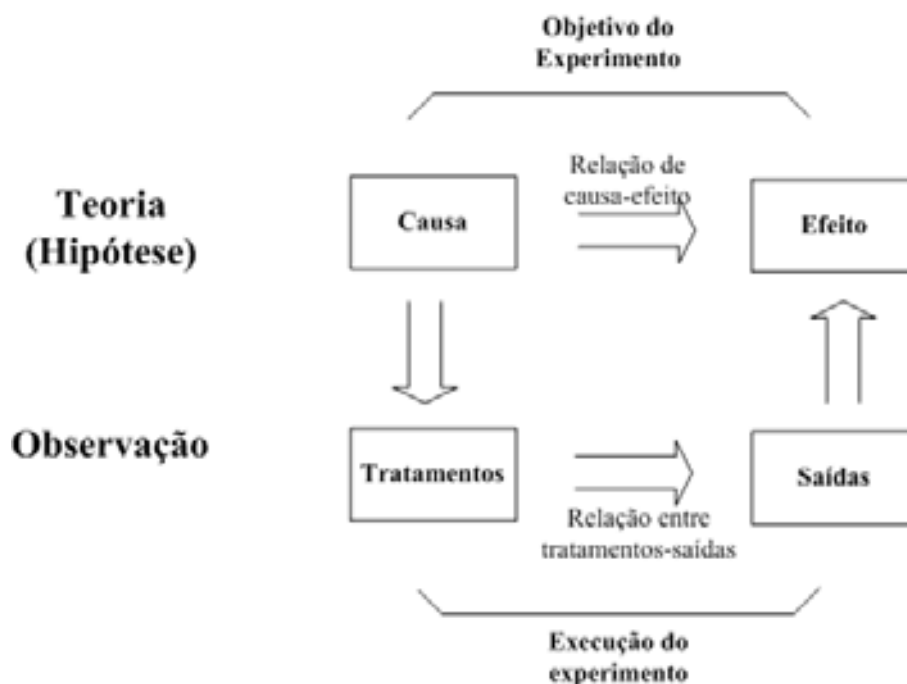
Essas implementações, juntamente com os testes preliminares, ajudaram a avaliar as potencialidades da solução e a viabilidade prática da solução nessas plataformas. Uma avaliação mais profunda e planejada dos principais aspectos da solução, no entanto, ainda se faz necessária para testar as hipóteses gerais formuladas no Capítulo 1. No próximo Capítulo, será apresentado um processo de experimentação desenvolvido para testar essas hipóteses e prover uma avaliação mais profunda dos principais aspectos da solução proposta, como, por exemplo o nível de compreensão e aceitação pelos usuários surdos, a qualidade da tradução, a viabilidade para traduções em tempo real, a produtividade no desenvolvimento das construções lingüísticas, dentre outros.

Capítulo 6

Experimentos

Segundo Wohlin et al. (2000), experimentos são investigações formais, rigorosas e controladas cujo principal objetivo é testar teorias ou hipóteses. Eles são, em geral, executados em ambientes controlados como, por exemplo, em laboratórios, e manipulam uma ou mais variáveis desse ambiente mantendo todas as outras em um nível fixo. A Figura 6.1, adaptada de Wohlin et al. (2000), apresenta os princípios básicos de um experimento.

Figura 6.1: Princípios básicos de um experimento



De acordo com a Figura 6.1, dada uma hipótese ou teoria que representa uma possível relação entre uma causa e um efeito, o experimento tenta obter conclusões sobre essa hipótese a partir de um conjunto de observações. Nessa observação, a causa definida

na teoria é associada a um conjunto de tratamentos (valores que as variáveis estudadas podem assumir) e o experimento tenta gerar um conjunto de saídas. Se o experimento tiver sido bem planejado e as saídas observadas estiverem associadas ao efeito definido na hipótese, é possível concluir que existe uma relação entre a causa e o efeito definido na hipótese formulada.

Neste capítulo, será apresentado um experimento conduzido com o objetivo de avaliar a solução proposta. A motivação do experimento é avaliar a solução proposta em três partes definidas de acordo com as hipóteses gerais formuladas no Capítulo 1. Na primeira parte, os conteúdos acessíveis gerados pela solução proposta são avaliados por usuários surdos com relação ao nível de compreensão e a qualidade de tradução desses conteúdos. Com isso, é possível investigar se os usuários são capazes de compreender os conteúdos acessíveis gerados e se sentem confortáveis com os mesmos e, por consequência, se é possível reduzir as suas barreiras de acesso à informação. Na segunda parte do experimento, o desempenho de tradução da solução proposta é avaliado, com o objetivo de investigar se a solução proposta era capaz de traduzir esses conteúdos em cenários que exigem tradução em tempo real. Por fim, na terceira parte, a ferramenta WikiLIBRAS é avaliada por usuários surdos, com o objetivo de investigar a produtividade dos usuários na geração das construções lingüísticas da solução proposta.

O processo do experimento envolveu a definição, planejamento e execução desse experimento, além da análise estatística e interpretação dos resultados. As etapas de definição e planejamento do experimento são apresentadas no Apêndice D e envolvem a definição do propósito do experimento e de seu contexto, além da formulação das hipóteses, definição das variáveis, sujeitos, objetos e instrumentos do experimento. As etapas de execução do experimento e de análise e interpretação dos resultados serão apresentadas nas próximas seções.

6.1 Execução do Experimento e Análise dos Dados

6.1.1 Avaliação do Nível de Compreensão dos Conteúdos

O experimento para avaliação do nível de compreensão dos conteúdos foi realizado nos dias 29 de Julho e 02 de Agosto de 2012 na Fundação Centro Integrado de Apoio à Pessoa com Deficiência (Funad) da Paraíba. O experimento contou com a participação de vinte usuários surdos selecionados por conveniência dos cursos de formação em LIBRAS da Funad. Esses usuários foram divididos aleatoriamente em dois grupos de dez usuários: um grupo para avaliar os objetos (conteúdos multimídia apresentados na Seção D.2.1)

com legendas e o outro grupo para avaliar os objetos com trilhas de LIBRAS gerados pela solução proposta. Durante todo o experimento, intérpretes de LIBRAS participaram do processo, intermediando a comunicação com os usuários.

Inicialmente, os usuários foram preparados para o teste. Na fase de preparação, o objetivo do teste foi apresentado para os usuários (avaliar o nível de compreensão em um conjunto de conteúdos multimídia), mas as hipóteses a serem testadas (definidas no Apêndice D.2.1) não foi apresentada¹. Em seguida, os usuários, já divididos em dois grupos, receberam uma cópia do questionário apresentado no Apêndice E, e foram instruídos a preencher a primeira parte dele com suas informações pessoais.

O grupo de usuários era heterogêneo e composto por onze mulheres e nove homens na faixa etária de 13 a 56 anos, e com uma média de idade de 28,6 anos. O grau de escolaridade e o nível de conhecimento em Português e LIBRAS dos usuários também foi observado e é apresentado na Figura 6.2 e na Tabela 6.1, respectivamente.

Figura 6.2: Grau de Escolaridade dos usuários do Teste



Após o preenchimento da primeira parte do questionário, os usuários (divididos em dois grupos) foram convidados a assistir os conteúdos multimídia do teste com o seu tratamento (legendas ou trilhas de LIBRAS geradas pela solução proposta). Cada conteúdo (vídeo) foi apresentado duas vezes aos usuários e após a apresentação de cada um desses conteúdos, os usuários foram convidados a responder as questões relacionadas ao con-

¹Para evitar que a hipótese a ser testada influenciasse a avaliação dos usuários, ela foi apresentada aos usuários apenas no final do experimento, após o preenchimento dos questionários

Tabela 6.1: Nível de conhecimento em Português e LIBRAS dos Usuários

Nível de Conhecimento	Valor Médio	Desvio Padrão
Português (1 a 6)	4,00	1,80
LIBRAS (1 a 6)	5,67	1,54%

teúdo apresentado (segunda parte do questionário). Por fim, os usuários foram convidados a responder as perguntas da terceira parte do questionário, avaliando subjetivamente alguns aspectos da solução. Algumas fotos deste experimento são apresentados na Figura 6.3.

Os resultados dos testes de compreensão de conteúdos deste experimento são apresentados na Figura 6.4. Conforme pode ser observado nessa figura, para todos os conteúdos avaliados, os usuários que assistiram conteúdos com legendas tiveram uma média de acertos menor que os usuários que assistiram conteúdos com trilhas de LIBRAS geradas pela solução proposta. Considerando todos os conteúdos, a taxa média de acertos para os usuários que assistiram vídeos com trilhas de LIBRAS (solução proposta) foi de 79,38% com um desvio padrão de 9,34%, enquanto que a taxa média de acertos para os usuários que assistiram vídeos com legendas foi de 25,63% com um desvio padrão de 19,86%.

Para uma melhor compreensão desses dados, um gráfico de caixa (ou *box plot*)² foi gerado com os resultados desses testes. Esse gráfico é apresentado na Figura 6.5. É importante ressaltar que os bigodes (*whiskers*) representados desse gráfico foram definidos de acordo com a proposta de Frigge et al. (1989), que definem os bigodes superior e inferior como o comprimento da caixa (i.e., a diferença entre o terceiro e o primeiro quartil) multiplicado por 1,5 e somado e subtraído terceiro (bigode superior) e primeiro quartil (bigode inferior), respectivamente.

De acordo com a Figura 6.5, é possível observar que existe um padrão claro de que os usuários que assistiram os conteúdos com trilhas de LIBRAS baseados na solução proposta obtiveram um melhor desempenho nos testes de compreensão de conteúdos do que os usuários que avaliaram os conteúdos com legendas. Para os conteúdos baseados na solução proposta, os valores da mediana, do primeiro e terceiro quartil da distribuição foram de 75,00%, 75,00% e 85,94%, respectivamente. Isso significa que 50% dos usuários obtiveram uma taxa de acertos na faixa de 75,00% a 85,94%. Além disso, nenhum *outlier*

²Os gráficos de *box plot* são uma forma de representar resumidamente a dispersão de um conjunto de amostras. Eles representam graficamente a mediana, primeiro e terceiro quartis da distribuição das amostras e são muito úteis para identificar *outliers* (as amostras que estão numericamente distantes do resto dos dados).

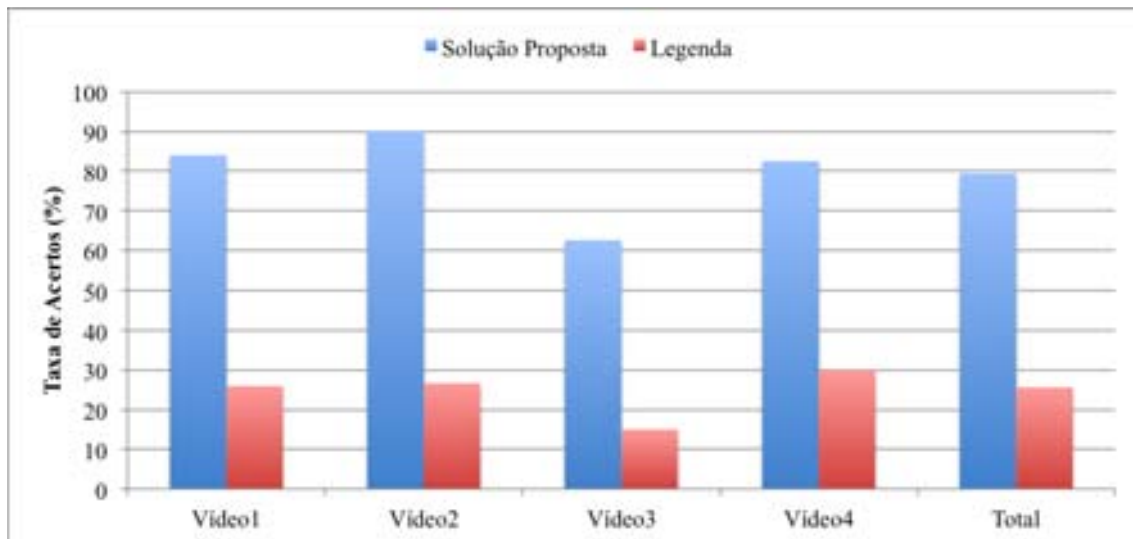
Figura 6.3: Algumas fotos do experimento com usuários (nível de compreensão) (a) Alguns usuários do experimento, (b) Fase de preparação dos usuários, (c) Usuários avaliando conteúdos legendados, (d) Usuários avaliando conteúdos com trilhas de LIBRAS geradas pela solução proposta.



foi identificado na distribuição, o que indica que todos os usuários obtiveram uma taxa média de acertos entre 68,75% e 93,75%.

Para os conteúdos com legendas, os valores da mediana, do primeiro e terceiro quartil da distribuição foram, respectivamente 21,88%, 18,75% e 25,00%. Isso significa que menos do que um quarto dos usuários obtiveram uma taxa de acertos superior a 25,00% das questões do experimento. Além disso, três *outliers* foram identificados nesta amostra. Um *outlier* negativo, representando um usuário que não obteve nenhum acerto (0,00%) e dois *outliers* positivos, representando dois usuários que obtiveram uma taxa de acertos de 37,5% e 75,00% para esses conteúdos. Como os *outliers*, neste caso, representam casos que podem realmente ocorrer na prática (isto é, usuários surdos que não sabem ler absolutamente nada - *outlier* negativo - e usuários surdos que sabem ler um pouco melhor - *outlier* positivo com 75% de acertos), essas amostras não foram removidas da análise.

Figura 6.4: Resultados dos testes de compreensão dos conteúdos

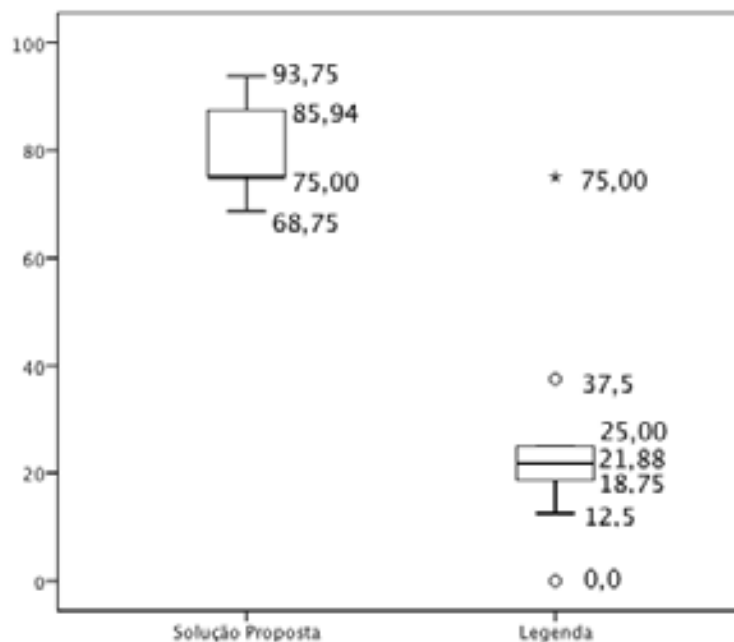


Para avaliar se essa diferença de desempenho é estatisticamente representativa, e, conseqüentemente, testar a hipótese relacionada ao nível de compreensão dos conteúdos (Hipótese 1 definida na Seção D.2.1), um teste-t foi então aplicado sobre todo o conjunto de amostras considerando um intervalo de confiança de 95% e 18 graus de liberdade (número de usuários ou amostras menos dois, ou seja, $20-2=18$). Os resultados da aplicação desse teste são apresentados na Tabela 6.4.

Tabela 6.2: Resultados do teste-t para a Hipótese 1 (compreensão dos conteúdos)

$\mu_{compreensao}(SP)$	79,38%	$\sigma_{compreensao}(SP)$	9,34%
$\mu_{compreensao}(Leg)$	25,63%	$\sigma_{compreensao}(Leg)$	19,86%
Graus de Confiança (p -valor)	0,95	Graus de Liberdade	18
t -valor	7,74	$t_{0,0025,18}$	2,12

De acordo com a Tabela 6.4, uma vez que o t -valor obtido para o teste foi maior do que o valor crítico para o teste-t com intervalo de confiança de 95% e 18 graus de liberdade ($t_{0,0025,18}$) [Wohlin et al. 2000], é possível concluir que a hipótese H_0 pode ser rejeitada. Dessa forma, é possível afirmar com um grau de confiança de 95% que existe uma diferença significativa de compreensão dos conteúdos quando usuários surdos são submetidos a conteúdos com trilhas de LIBRAS baseadas na solução proposta com relação ao nível de compreensão quando os usuários são submetidos a conteúdos com

Figura 6.5: Gráfico de caixa (*box plot*) para os testes de compreensão dos conteúdos

legendas.

Por fim, alguns aspectos da solução como, por exemplo, a qualidade de tradução, naturalidade da apresentação, dentre outros, também foram avaliados de forma subjetiva pelos usuários. Nessa avaliação, após assistir todos os conteúdos, os usuários atribuíram conceitos de 1 a 6 para cada um desses aspectos. Alguns deles, no entanto, que não fazem sentido para os conteúdos com legendas, como, por exemplo, naturalidade da apresentação, qualidade dos movimentos das mãos e qualidade das expressões faciais, e não foram avaliados pelos usuários do grupo de conteúdos com legendas. Os resultados dessa avaliação são apresentados na Tabela 6.3.

De acordo com a Tabela 6.3, o nível de compreensão e a qualidade da tradução, em média, obtiveram valores médios moderados (4,60 e 4,60, respectivamente) nos conteúdos gerados pela solução proposta. Além disso, esses valores foram maiores do que os valores correspondentes nos conteúdos com legendas (3,70 e 4,13, respectivamente). Os altos valores de desvio padrão para os dois tipos de conteúdos, no entanto, mostram que as opiniões dos usuários divergiram com relação a avaliação desses aspectos.

Um aspecto importante a ser observado, no entanto, é que a avaliação do aspecto "nível de compreensão" para os conteúdos com legendas não foi compatível com os testes de compreensão realizados com os usuários. Embora os usuários que avaliaram estes

Tabela 6.3: Valores médios para os aspectos avaliados (escalas de 1 a 6)

Aspecto avaliado	Solução Proposta		Legenda	
	Valor médio	Desvio Padrão	Valor Médio	Desvio Padrão
Nível de Compreensão	4,60	1,68	3,70	2,33
Fluência Gramatical	4,60	1,56	4,13	2,05
Naturalidade	4,40	1,74	-	-
Qualidade dos movimentos	4,8	1,4	-	-
Qualidade das expressões faciais	4,56	1,89	-	-

conteúdos tenham indicado que eles foram razoavelmente compreendidos (3,70), os resultados dos testes de compreensão mostraram que a maioria dos usuários não haviam compreendido bem os conteúdos (taxa média de acerto de 25% das questões).

É possível verificar essa incompatibilidade, analisando a correlação entre as duas variáveis: (1) resultados dos testes de compreensão e (2) avaliação do "nível de compreensão". Ao calcularmos os valores do coeficiente de correlação de Pearson e do coeficiente de correlação de postos de Spearman [Wohlin et al. 2000] para essas variáveis, os valores obtidos foram de 0,033 e -0,182, respectivamente, o que indica uma baixa correlação entre as variáveis. Uma das possíveis explicações para isso, é que segundo Wohlin et al. (2000), os humanos têm medo de serem avaliados e quando passam por um processo de avaliação tentam parecer melhor do que realmente são, o que pode atrapalhar a saída do experimento.

Além disso, com relação a fluência gramatical, outra inconsistência nos resultados é que muitos usuários indicaram que os conteúdos com legendas eram compatíveis com a gramática de LIBRAS (4,13 de valor médio), enquanto, na realidade, os conteúdos estavam sendo transmitidos na gramática da língua portuguesa.

A naturalidade e qualidade da apresentação também foram avaliados. A naturalidade da apresentação também obteve um valor médio moderado (4,40), mas foi o aspecto avaliado que obteve o pior valor médio. Esse resultado é compatível e próximo dos valores obtidos para a qualidade dos movimentos das mãos (4,80) e de qualidade das expressões faciais (4,56). Contudo, de forma similar, as outras medidas, os valores de desvio padrão também foram altos, o que sugere que existem alguns indícios de divergência entre os usuários.

Assim como em San-segundo et al. (2011), algumas prováveis causas para essas divergências foram observadas durante este experimento. Por exemplo, durante o experimento,

os usuários divergiram sobre a estrutura de algumas sentenças em LIBRAS. De forma similar as outras línguas de sinais (como por exemplo, na LSE [San-segundo et al. 2011]), a LIBRAS possui um alto grau de flexibilidade na estruturação das sentenças. Essa flexibilidade é algumas vezes não é bem compreendida e algumas das possibilidades foram consideradas como sentenças incorretas. Além disso, alguns usuários também divergiram com relação a correta sinalização de alguns sinais, como, por exemplo, os sinais CAFÉ e MERCADO.

Uma alternativa para reduzir essas divergências no LibrasTV, por exemplo, seria customizar os Dicionários de LIBRAS nos receptores de TV dos usuários. O desenvolvimento desses dicionários customizados, no entanto, é uma tarefa muito trabalhosa. Outra alternativa seria investir mais esforços para padronizar a LIBRAS. Nesse caso, uma maior disseminação de LIBRAS nas TIC poderia auxiliar nesse processo de padronização, como também vem ocorrendo em outras línguas menores na Espanha [San-segundo et al. 2011].

Em consequência dessa forte divergência, ao aplicarmos o teste-t (ver Tabela 6.4), não é possível rejeitar a segunda hipótese deste experimento relacionada a adequação a gramática de LIBRAS (Hipótese 2 definida na Seção D.2.1) e, conseqüentemente, não é possível afirmar que os conteúdos gerados pela solução proposta se adéquam melhor a gramática de LIBRAS do que os conteúdos com legendas. Os bons resultados nos testes de compreensão dos conteúdos, no entanto, são um indício de que a qualidade de tradução é adequada. Uma análise qualitativa mais profunda, contudo, precisa ser desenvolvida para avaliar esse aspecto, sendo uma das propostas de trabalho futuro.

Tabela 6.4: Resultados do teste-t para a Hipótese 2 (adequação a gramática de LIBRAS)

$\mu_{traducao}(SP)$	4,60	$\sigma_{traducao}(SP)$	1,56
$\mu_{traducao}(Leg)$	4,13	$\sigma_{traducao}(Leg)$	2,05
Grau de Confiança (p -valor)	0,95	Graus de Liberdade	18
t -valor	0,58	$t_{0,0025,18}$	2,12

6.1.2 Avaliação do Desempenho de Tradução

O experimento para avaliar o desempenho de tradução foi realizado utilizando testes objetivos com um sinal real de TV Digital durante um período de 24 horas (das 9 horas do dia 20 de Dezembro de 2011 até as 09 horas do dia 21 de Dezembro de 2011). Durante esse período, o sinal da "TV Correio", uma emissora de TV Digital Brasileira afiliada

da "TV Record", foi sintonizado pelo "Retransmissor IP MPEG-2 TS" e o fluxo MPEG-2 TS desse sinal foi transmitido para o "Tradutor de LIBRAS" e para o Multiplexador. Sempre que pacotes de legendas eram recebidos pelo "Tradutor de LIBRAS", o processo de tradução e geração das trilhas de LIBRAS era executado pelo protótipo e os atrasos de cada componente eram medidos e armazenados, resultando na tradução de cerca de 2200 sentenças.

Na Tabela 6.5 são apresentados os valores dos atrasos médio, máximo, mínimo e o desvio padrão do experimento para cada um dos componentes. De acordo com a Tabela 6.5, o atraso médio para executar todos os componentes do LibrasTV foi menor do que 44 milissegundos. O atraso máximo obtido (considerando o atraso máximo de cada um dos componentes) foi de 142 milissegundos, enquanto que o atraso mínimo obtido foi de 20,51 milissegundos.

Tabela 6.5: Medidas do desempenho de tradução para os componentes do protótipo LibrasTV

Componentes	Valor Médio (ms)	Desvio Padrão (ms)	Valor Máximo (ms)	Valor Mínimo (ms)
Filtragem e Extração de Legendas	0,024	0,022	0,554	0,017
Tradução Automática	0,975	2,957	80,126	0,220
Codificação	0,215	0,089	1,061	0,072
Decodificação	0,170	0,143	0,519	0,020
Animacão e Distribuição	42,445	8,747	59,998	20,000
Total	43,805	9,434	142,21	20,509

Para avaliar se esse atraso de tradução atende o critério de tradução em tempo real definido na Seção D.2.2 (isto é, $\mu_{atraso}(SP) \leq 4$ segundos), um teste-t foi aplicado sobre essa amostra considerando um grau de confiança de 95% e 2192 graus de liberdade (número de amostras menos dois, ou seja, $2194-2=2192$). Os resultados da aplicação desse teste são apresentados na Tabela 6.6.

De acordo com a Tabela 6.6, uma vez que t -valor obtido para o teste foi maior do que o valor crítico para o teste-t com intervalo de confiança de 95% e 2192 graus de liberdade ($t_{0,0025,2192}$) [Wohlin et al. 2000], é possível concluir que a hipótese H_0 dessa parte do experimento pode ser rejeitada (Hipótese 1 definida na Seção D.2.2). Dessa forma, considerando que esse experimento foi executado com um vocabulário aberto e representativo e, em um cenário real, é possível afirmar com um grau de confiança de

Tabela 6.6: Resultados do teste-t para a Hipótese desse experimento (desempenho ou atraso da tradução)

$\mu_{atraso}(SP)$	43,805 miliseg.	$\sigma_{atraso}(SP)$	9,434 miliseg.
Atraso de referência	4 seg.		
fGrau de Confiança (p -valor)	0,95	Graus de Liberdade	2192
t -valor	19.632,8	$t_{0,0025,2192}$	1,96

95% que a solução proposta é capaz de traduzir conteúdos em tempo real.

Além disso, o tempo (atraso) médio de tradução obtido é provavelmente bem inferior ao tempo gasto por uma tradução humana, embora com uma qualidade de tradução inferior também.

6.1.3 Avaliação da Ferramenta de Computação Humana

O experimento para avaliação da ferramenta de computação humana (WikiLIBRAS) também foi realizado nos dias 29 de Julho e 02 de Agosto de 2012 na Funad-PB. Ele contou com a participação de onze usuários surdos e três intérpretes de LIBRAS selecionados por conveniência dos cursos de formação em LIBRAS da Funad-PB.

Inicialmente, os usuários foram preparados de forma similar ao experimento de avaliação do nível de compreensão dos conteúdos (ver Seção 6.1.1) e preencheram a primeira parte do questionário (ver Apêndice E) com suas informações pessoais.

O grupo de usuários também era heterogêneo e composto por sete mulheres e sete homens na faixa etária de 12 a 42 anos, e com uma média de idade de 25,4 anos. O grau de escolaridade e o nível de conhecimento em Português e LIBRAS dos usuários também foi observado e é apresentado na Figura 6.6 e na Tabela 6.7, respectivamente.

Tabela 6.7: Nível de conhecimento em Português e LIBRAS dos Usuários

Nível de Conhecimento	Valor Médio	Desvio Padrão
Português (1 a 6)	4,43	1,24
LIBRAS (1 a 6)	5,07	1,16%

Os usuários foram então convidados a gerar os sinais de LIBRAS propostos (ver Seção D.2.3) utilizando o WikiLIBRAS e a preencher a segunda e a terceira parte do questionário. Durante o experimento, o tempo médio gasto pelos usuários para gerar os sinais

Figura 6.6: Grau de Escolaridade dos usuários do Teste



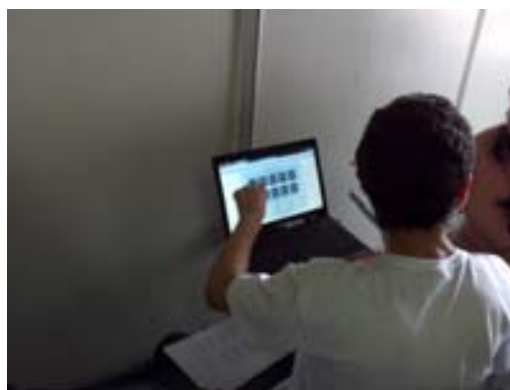
(eficiência) também foi armazenado no WikiLIBRAS. Na Figura 6.7 são apresentadas algumas fotos do experimento realizado com surdos e intérpretes de LIBRAS na Funad-PB.

Figura 6.7: Algumas fotos do experimento com usuários (WikiLIBRAS).

(a)



(b)



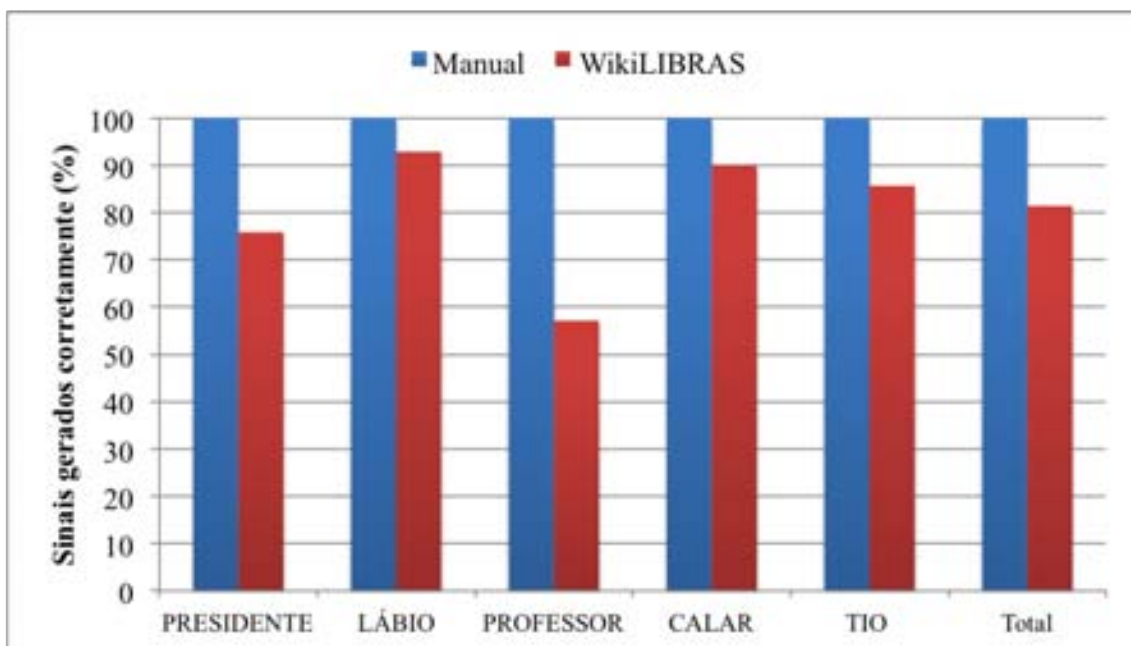
Para confrontar com o desempenho dos usuários no WikiLIBRAS, um experimento similar também foi executado com *designers-3D* no LAViD-UFPB. Nesse experimento, três experientes *designers-3D* foram convidados a gerar o mesmo conjunto de sinais utilizando a ferramenta de modelagem e animação Blender³. Para auxiliar os *designers-3D* na

³www.blender.org/

animação dos sinais, um vídeo com cada sinal representado por um intérprete de LIBRAS foi gerado e passado para os *designers* para ser usado como referência na geração dos sinais. Os três *designers* que participaram do experimento são experientes e participam de projetos de pesquisa no LAViD envolvendo modelagem e animação-3D, sendo dois deles alunos de graduação e o outro aluno de pós-graduação da UFPB.

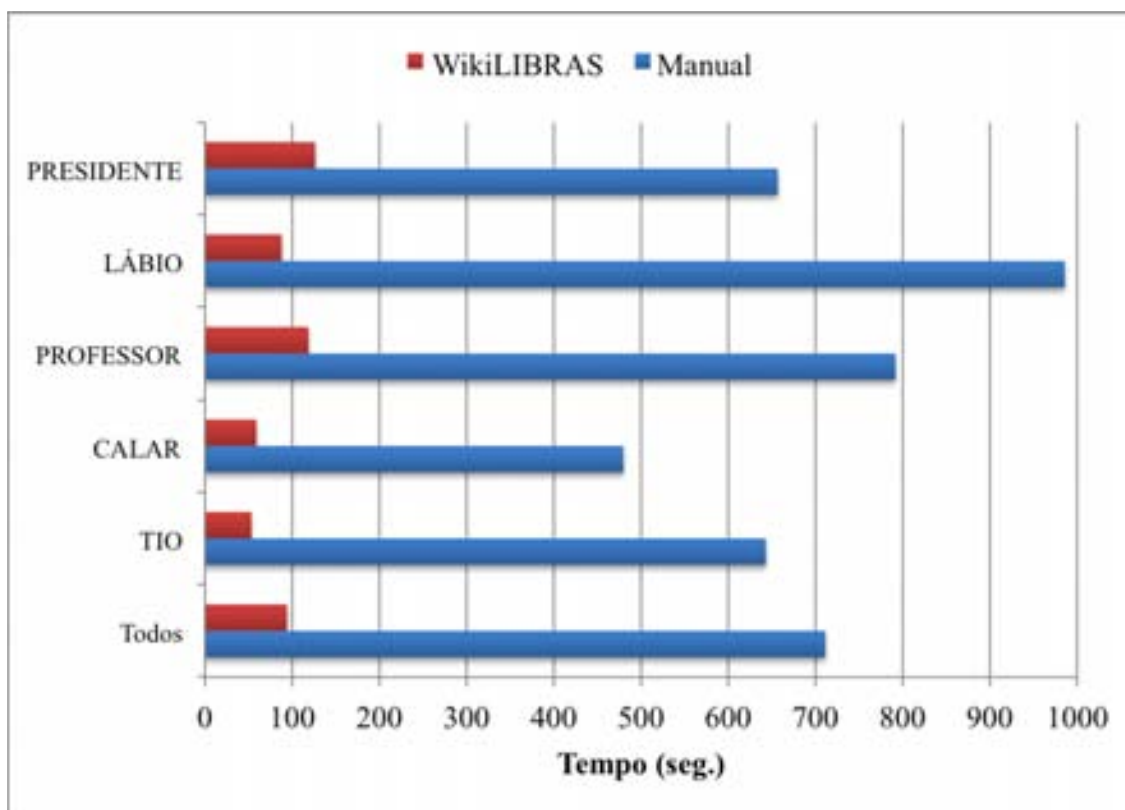
Os principais resultados desses dois experimentos são apresentados nas Figuras 6.8 e 6.9. De acordo com a Figura 6.8, é possível observar que os *designers*-3D conseguiram gerar todos os sinais corretamente na ferramenta de animação, enquanto que os surdos e intérpretes de LIBRAS conseguiram gerar, em média, cerca de 81,43% dos sinais no WikiLIBRAS. Com relação ao tempo médio para gerar os sinais corretamente (ver Figura 6.9), no entanto, o tempo médio gasto pelos usuários no WikiLIBRAS (93,96 segundos) foi bem inferior ao tempo médio gasto pelos *designers* na ferramenta de animação (711,33 segundos).

Figura 6.8: Taxa de sinais gerados corretamente pelos usuários (WikiLIBRAS vs Manual).



Analisando a geração de cada sinal individualmente, é possível observar que os usuários tiveram, em média, mais dificuldades para gerar os sinais PROFESSOR e PRESIDENTE no WikiLIBRAS. O sinal PROFESSOR foi gerado corretamente por 57,14% dos usuários, enquanto que o sinal PRESIDENTE foi gerado corretamente por 75,86% dos

Figura 6.9: Tempo médio para gerar os sinais corretamente (WikiLIBRAS vs Manual).



usuários. Para os outros sinais (LÁBIO, CALAR e TIO), os usuários encontraram menos dificuldades e obtiveram uma taxa média de acertos superior a 85%. Esses resultados são compatíveis com o tempo médio gasto pelos usuários para gerar os sinais. Nos sinais que os usuários tiveram mais dificuldades, como, por exemplo, nos sinais PROFESSOR e PRESIDENTE, o tempo médio para gerar os sinais (118,72 segundos e 126,33 segundos, respectivamente) foi maior do que o tempo médio necessário para gerar os sinais LÁBIO, CALAR e TIO (87,88 segundos, 58,91 segundos e 53,31 segundos, respectivamente).

Dentre as dificuldades apontadas pelos usuários para gerar os sinais PROFESSOR e PRESIDENTE, destaca-se a dificuldade de alguns usuários em entender alguns parâmetros utilizados na interface Web do WikiLIBRAS. Dessa forma, uma proposta de trabalho futuro é incluir nesta interface, vídeos com intérpretes de LIBRAS para auxiliar os usuários durante a navegação.

Para avaliar se esta diferença com relação ao tempo médio de geração dos sinais no WikiLIBRAS é estatisticamente representativa e, conseqüentemente, testar a hipótese dessa parte do experimento (Hipótese 1 definida na Seção D.2.3), um teste-t foi nova-

mente aplicado considerando um grau de confiança de 95% e 15 graus de liberdade (número de usuários menos dois, ou seja, $17-2=15$). Os resultados da aplicação desse teste são apresentados na Tabela 6.8.

Tabela 6.8: Resultados do teste-t para a Hipótese desse experimento (tempo médio para gerar os sinais)

$\mu_{tempo}(WL)$	93,96 seg.	$\sigma_{tempo}(WL)$	74,57 seg.
$\mu_{tempo}(MAN)$	711,33 seg.	$\sigma_{tempo}(MAN)$	260,33 seg.
Grau de Confiança (p -valor)	0,95	Graus de Liberdade	15
t -valor	12,53	$t_{0,0025,15}$	2,13

De acordo com a Tabela 6.8, uma vez que o t -valor obtido para o teste foi maior do que o valor crítico para o teste-t com intervalo de confiança de 95% e 15 graus de liberdade ($t_{0,0025,15}$) [Wohlin et al. 2000], é possível concluir que a hipótese H_0 pode ser rejeitada. Dessa forma, é possível afirmar com um grau de confiança de 95% que existe uma diferença significativa no tempo médio de geração dos sinais quando especialistas em LIBRAS utilizam o WikiLIBRAS, quando comparado com o tempo médio de geração desses mesmos sinais através de ferramentas de animação.

Além de produzir os conteúdos num intervalo de tempo significativamente menor, o número de surdos e intérpretes de LIBRAS que existe no Brasil é muito maior do que o número de *designers*-3D disponíveis, além do que os *designers*-3D necessitam de vídeos de referências dos sinais ou precisam aprender os sinais em LIBRAS. Dessa forma, é possível criar um Dicionário de LIBRAS de forma mais produtiva utilizando o WikiLIBRAS, especialmente considerando que um Dicionário de LIBRAS possui cerca de 6.000 sinais [Acesso Brasil 2012].

Por fim, a Tabela 6.8 apresenta a avaliação subjetiva dos principais aspectos do WikiLIBRAS por usuários surdos e intérpretes de LIBRAS. De acordo com a Tabela 6.9, a usabilidade do WikiLIBRAS obteve o maior valor médio (4,77) dentre os aspectos avaliados. Esse resultado é compatível com os valores médios de avaliação das telas de configuração de mão (4,69), expressão facial (4,92) e ponto de articulação (4,69) do WikiLIBRAS. A tela de orientação da palma da mão, no entanto, provavelmente causou alguma confusão entre os usuários e obteve um dos piores valores médios (3,92). Conforme mencionado anteriormente, uma das proposta de trabalho futuro para melhorar esse aspecto, é incluir na interface do WikiLIBRAS, vídeos com intérpretes de LIBRAS, para auxiliar os usuários durante a navegação.

Tabela 6.9: Valores médios para os aspectos avaliados (escalas de 1 a 6)

Questões	Valor Médio	Desvio Padrão
Usabilidade	4,77	1,59
Naturalidade da animação	3,62	1,71
Tela de Configuração de Mão	4,69	1,84
Tela de Orientação	3,92	1,93
Tela de Expressão Facial	4,92	1,04
Tela de Ponto de Articulação	4,69	1,55

A naturalidade da animação 3D obteve o menor valor médio (3,62) dentre os aspectos avaliados. Isso pode ser explicado porque a naturalidade da sinalização baseada em avatares ainda não é comparável com a sinalização humana, especialmente porque os sinais são gerados de forma semi-automática. Além disso, essa pouca naturalidade pode ter influenciado nos resultados dos testes de compreensão e de avaliação da adequação gramatical. Dessa forma, é necessário investir mais esforços para melhorar a flexibilidade e naturalidade do avatar-3D. Uma possível alternativa é integrar ferramentas de captura de movimento, como, por exemplo, o Microsoft Kinect, para capturar os movimentos dos sinais.

6.2 Considerações

Nesse capítulo foi apresentado um processo de experimentação para a solução proposta neste trabalho. Esse processo envolveu a definição, planejamento e execução de um experimento conduzido para testar as hipóteses gerais do trabalho apresentadas no Capítulo 1. Nesse processo, foram formuladas hipóteses mais específicas para avaliar o nível de compreensão de usuários surdos nos conteúdos gerados pela solução proposta, para avaliar o desempenho (tempo médio de tradução) e para avaliar a produtividade do WikiLIBRAS no desenvolvimento das construções lingüísticas da solução.

Os resultados desse experimento mostraram que três das quatro hipóteses formuladas puderam ser rejeitadas, indicando que a solução proposta é capaz de reduzir as barreiras de acesso a informação, uma vez que ela conseguiu melhorar o nível de compreensão dos conteúdos com relação aos conteúdos com legendas; é capaz de traduzir conteúdos também em cenários que exigem tradução em tempo real, como, por exemplo, na TV; e sua ferramenta de computação humana é capaz de reduzir o tempo médio de produ-

ção das suas construções lingüísticas, quando comparado com a geração manual dessas construções.

Adicionalmente, conforme evidenciado em trabalhos anteriores [Cox et al. 2002][Kipp et al. 2012][San-segundo et al. 2011], embora as soluções baseadas em avatares não sejam a primeira opção dos surdos que preferem tradução humana, a solução proposta é apresentada como uma alternativa prática e viável, capaz de tornar os conteúdos multimídia acessíveis para surdos, especialmente quando intérpretes humanos não estão disponíveis.

No próximo capítulo, serão apresentadas as conclusões sobre o presente trabalho, incluindo uma discussão sobre as principais contribuições e propostas de trabalhos futuros.

Capítulo 7

Conclusões

Neste trabalho, foi defendida a tese de que é possível gerar automaticamente trilhas (vídeos) de LIBRAS para os surdos em conteúdos digitais multimídia, a partir da tradução automática de fluxos de legendas. Na proposta, fluxos de legendas em língua portuguesa são traduzidos automaticamente (isto é, sem intervenção humana direta) para vídeos em LIBRAS. Esse vídeo de LIBRAS é então embutido no conteúdo digital multimídia original (através de mixagem, multiplexação ou transmissão para outros dispositivos), adicionando uma camada acessível nesse conteúdo e tornando-o, portanto, acessível para surdos. A proposta também prevê que a solução seja capaz gerar os conteúdos acessíveis em cenários que exijam tradução em tempo real (por exemplo, na TV) e que seja dotada de mecanismos que permitam a geração semi-automática e manutenção das suas construções lingüísticas de forma colaborativa.

Para testar as hipóteses formuladas a partir dessa tese (ver Capítulo 1), a arquitetura de um sistema de tradução automática de legendas em Português para LIBRAS foi inicialmente projetada e desenvolvida. Uma característica importante dessa arquitetura é que ela inclui uma estratégia eficiente de tradução automática de textos em língua portuguesa para glosa em LIBRAS e utiliza dicionários de LIBRAS que armazenam um conjunto de vídeo dos sinais, para minimizar a carga de processamento necessária para gerar o vídeo (trilha) de LIBRAS em tempo real.

Outro importante aspecto da solução é que ela possui uma infraestrutura que envolve uma ferramenta de computação humana, linguagens de descrição de regras e sinais e o modelo de um avatar-3D utilizados para o desenvolvimento e manutenção das construções lingüísticas da solução de forma eficiente e semi-automática.

Adicionalmente, foram desenvolvidos protótipos da solução proposta para as plataformas de TV Digital, Web e Cinema Digital, e um processo de experimentação bem definido, envolvendo usuários surdos brasileiros, foi conduzido para avaliar a solução e testar as hipóteses gerais do trabalho.

Os resultados desse processo de experimentação mostraram que a solução proposta é eficiente e capaz de gerar e embarcar as trilhas de LIBRAS em diferentes conteúdos e cenários, incluindo os cenários que exigem tradução em tempo real e que são de domínio geral (por exemplo, TV Digital). Além disso, a solução proposta conseguiu melhorar o nível de compreensão dos conteúdos com relação aos conteúdos com legendas, o que indica que ela é capaz de reduzir as barreiras de acesso à informação nos conteúdos multimídia. A ferramenta de computação humana que integra a solução também foi avaliada, e foi possível observar que ela é capaz de reduzir o tempo médio de produção das suas construções linguísticas, quando comparado com a geração manual dessas construções.

Por fim, conforme evidenciado em trabalhos anteriores [Cox et al. 2002][Kipp et al. 2012][San-segundo et al. 2011], embora as soluções baseadas em avatares não sejam a primeira opção dos surdos que preferem tradução humana, a solução proposta é apresentada como uma alternativa prática e viável, capaz de tornar os conteúdos multimídia acessíveis para surdos, especialmente quando intérpretes humanos não estão disponíveis.

Nas próximas seções, serão detalhadas e discutidas as principais contribuições e as principais perspectivas para este trabalho.

7.1 Principais Contribuições

Considerando os resultados obtidos durante o processo de desenvolvimento deste trabalho, pode-se observar que uma lista importante de contribuições foram produzidas. As contribuições científicas gerais, desenvolvidas pelo autor deste trabalho serão classificadas como "contribuições gerais", enquanto que as contribuições desenvolvidas em cooperação com outros pesquisadores do grupo de pesquisa serão classificadas como "contribuições colaterais". Por fim, uma produção técnica e científica diversificada também foi gerada durante o desenvolvimento deste trabalho, resultando em um conjunto de publicações em periódicos e conferências nacionais e internacionais, patentes registradas e projetos aprovados. Todas essas contribuições serão apresentadas nas próximas subseções.

7.1.1 Contribuições Científicas Gerais

As contribuições científicas gerais deste trabalho contemplam a arquitetura geral da solução proposta nesse trabalho, a estratégia de tradução automática de português para glosa, a ferramenta de computação humana utilizada para o desenvolvimento das construções linguísticas da solução e o protocolo de codificação de LIBRAS desenvolvido para

sistemas de TV Digital. Essas contribuições são listadas e detalhadas abaixo:

- **Definição e projeto de uma arquitetura para geração automática e em tempo real de trilhas de LIBRAS a partir de fluxos de legendas em língua portuguesa.**

Essa arquitetura, base da solução proposta, é composta por um conjunto de componentes de software (Filtragem, Extração de Legendas, Tradução Automática, Animação, Sincronização e Distribuição) que viabilizam a geração dessas trilhas a partir da tradução automática dos fluxos de legendas em língua portuguesa para LIBRAS e embutem essas trilhas de LIBRAS no conteúdo multimídia utilizando uma estratégia de mixagem, multiplexação ou transmissão para outro *display*, tornando os conteúdos multimídia acessíveis para surdos.

Conforme mencionado, uma característica importante e inédita dessa solução é que ela inclui uma estratégia eficiente de tradução automática de textos em língua portuguesa para glosas em LIBRAS e utiliza dicionários de LIBRAS para minimizar os recursos computacionais necessários para gerar os conteúdos de LIBRAS em tempo real. Além disso, dicionários de LIBRAS são utilizados para evitar a renderização dos sinais em tempo real, uma vez que essa tarefa consome muito tempo. Esses dicionários armazenam vídeos dos sinais de LIBRAS pré-renderizados e cada sinal possui um código (por exemplo, sua representação textual em glosa) associado com esse vídeo. Dessa forma é possível gerar um vídeo de LIBRAS a partir da combinação de sinais no dicionário de LIBRAS.

Outro importante e inédito aspecto dessa arquitetura é ela dispõe de uma infraestrutura para geração semi-automática das construções lingüísticas (sinais e regras de tradução) da solução.

- **Projeto e desenvolvimento de uma estratégia eficiente de tradução automática de textos em Português para glosa em LIBRAS.**

Na arquitetura da solução proposta, uma estratégia de tradução eficiente de textos em Português para glosa em LIBRAS foi desenvolvida (componente de Tradução Automática). Essa estratégia combina métodos de compressão estatística utilizados para classificar os *tokens* (palavras) de entrada, estratégias de simplificação textual para reduzir a complexidade do texto de entrada e um conjunto de regras morfológicas e sintáticas, definido por especialistas.

- **Projeto e desenvolvimento de uma ferramenta de computação humana que permite a geração e manutenção dos sinais e regras de tradução.**

Uma ferramenta de computação humana também foi projetada e desenvolvida. Essa ferramenta, juntamente com as linguagem de descrição de sinais e regras e o modelo

do avatar-3D, compõe a infra-estrutura que permite que especialistas e usuários surdos colaborem na geração e manutenção de um banco de sinais em LIBRAS (Dicionário de LIBRAS) e uma base de dados de regras de tradução.

- **Proposta e especificação de um protocolo de codificação de LIBRAS para Sistemas de TV Digital.**

Para viabilizar a transmissão de legendas em LIBRAS no fluxo de TV Digital utilizando pouca largura de banda e excluindo a necessidade de modificações no *middleware* do sistema de TV Digital, um protocolo de alto nível de abstração também foi proposto e especificado para codificar as seqüências de glosas em LIBRAS no fluxo de TV Digital. Esse protocolo é composto basicamente por dois tipos de mensagens: a LIBRASControlMessage (LCM), uma mensagem de controle, utilizada para transmitir informações relacionadas ao tamanho e posição inicial da janela de LIBRAS; e a LIBRASDataMessage (LDM), uma mensagem de dados, utilizada para transmitir as glosas geradas pelo componente de Tradução Automática. Essas mensagens são então envelopadas em eventos DSM-CC para serem embutidas no fluxo TS da TV, tornando possível a sincronização da aplicação com outras mídias relacionadas como, por exemplo, o áudio e o vídeo ou a legenda do programa. As mensagens LCM e LDM são encapsuladas no campo de dados privados (*privateDataBytes*) dos eventos DSM-CC e as informações de sincronização (*timestamps*) são encapsuladas no campo de referência temporal (*eventNPT*).

Essa proposta de protocolo encontra-se atualmente em discussão no Fórum SBTVD e é uma das propostas candidatas a se tornar padrão de transmissão de LIBRAS no SBTVD.

7.1.2 Contribuições Colaterais

Dentre as contribuições colaterais, é possível destacar o modelo do avatar-3D humanóide, as linguagens formais de descrição de sinais e regras de tradução e as implementações dos protótipos LibrasTV, LibrasWeb e CineLIBRAS. Essas contribuições são apresentadas e detalhadas abaixo:

- **Modelagem de um agente animado virtual 3D humanóide.**

Para representar as animações (sinais) utilizados pela solução proposta na geração do vídeo de LIBRAS, um agente animado virtual 3D (um avatar-3D) foi modelado e implementado no *software* Blender¹. Esse modelo contém uma armadura de 82

¹www.blender.org/

ossos, com ossos distribuídos nas mãos, braços, face e corpo, e foi projetado para representar todos os possíveis fonemas definidos na LIBRAS (configuração de mão, pontos de articulação, orientação, movimento e expressões faciais).

- **Especificação de uma linguagem formal para descrição de sinais em LIBRAS.**
Uma linguagem formal para descrever formalmente sinais em LIBRAS também foi especificada. Essa linguagem permite que os parâmetros que compõem um sinal sejam descritos por colaboradores e, a partir desses parâmetros, animações podem ser geradas com base no modelo de um agente animado virtual-3D desenvolvido.
- **Especificação de uma linguagem para descrição de regras de tradução.**
Uma linguagem também foi definida para descrever formalmente as regras de tradução. De forma similar a linguagem de descrição, a partir dessa linguagem é possível que especialistas e colaboradores auxiliem no desenvolvimento dos tradutores automáticos para LIBRAS, descrevendo regras de tradução.
- **Implementação de protótipos da solução proposta para plataformas como TV Digital, Web e Cinema Digital.**
Protótipos da solução proposta foram desenvolvidos para as plataformas de TV Digital, Web e Cinema Digital. O desenvolvimento desse protótipos permitiu a construção de uma prova de conceito da solução nessas plataformas, auxiliando na redução das barreiras de comunicação dos usuários surdos. As arquiteturas e adaptações da solução proposta para essas plataformas também são contribuições deste trabalho.

7.1.3 Produção Técnica-Científica

Por fim, uma produção técnica e científica também foi gerada, incluindo publicações em periódicos e conferências nacionais e internacionais, patentes registradas e projetos aprovados. O conjunto dessas produções é detalhado abaixo:

- **Três Artigos Aceitos para Publicação em Periódicos:**
 1. **Araújo, T. M. U., G. L. Souza Filho et al. (2012)** 'Accessibility as a Service: Augmenting Multimedia Content with Sign Language Video Tracks'. *Journal of Research and Practice in Information Technology* (Aceito para Publicação).
 2. **Araújo, T. M. U., G. L. Souza Filho, et al. (2012)**, 'Automatic Generation of Brazilian Sign Language Windows for Digital TV Systems', *Journal of the Brazilian Computer Society* (Aceito para Publicação).
 3. **Ferreira, F. L. S., T. M. U. Araújo, G. L. Souza Filho, et al. (2012)**, 'Generating Window of Sign Languages on ITU J.200-Based Middlewares', *In-*

ternational Journal of Multimedia Data Engineering and Management **3**(2), 20-40.

- **Um Artigo Submetido para Periódico (por convite):**

1. Silva, D. A. N. S., **T. M. U. Araújo**, G. L. Souza Filho et al. (2012), 'A Formal Language to Describe and Animate Signs in Brazilian Sign Language'. *SBC Journal on 3D Interactive Systems*

- **Dois Prêmios de Melhor Artigo em Conferência Nacional e Internacional:**

1. Best paper award runner-up no IEEE International Symposium on Multimedia (ISM2011), IEEE International Symposium on Multimedia (ISM2011).
2. Artigo selecionado entre os "best papers" no XIV Simpósio de Realidade Virtual e Aumentada, SVR 2012 (SVR2012).

- **Cinco Artigos Completos Publicados em Conferências Nacionais e Internacionais:**

1. Silva, D.A.N.S., **T. M. U. Araújo**, G. L. Souza Filho, et al. (2012), FleX-LIBRAS: Description and Animation of Signs in Brazilian Sign Language, em 'Anais do XIV Simpósio de Realidade Virtual e Aumentada, SVR 2012', Niterói, Brasil, pp. 1-10.
2. Ferreira, F. L. S., **T. M. U. Araújo**, G. L. Souza Filho, et al. (2011), Providing Support for Sign Languages in Middlewares Compliant with ITU J.202, em 'Proceedings of the 2011 IEEE International Symposium of Multimedia, ISM'11', Dana Point, EUA, pp. 149-156.
3. Lemos, F. H., **T. M. U. Araújo**, G. L. Souza Filho, et al. (2011), Uma Proposta de Protocolo de Codificação de LIBRAS para Sistemas de TV Digital, em 'Anais do XVII Simpósio Brasileiro de Sistemas Multimídia e Web, Webmedia'11', Florianópolis, Brasil, pp. 103-110.
4. Silva, D.A.N.S., **T. M. U. Araújo**, G. L. Souza Filho, et al. (2011), Wiki-LIBRAS: Construção Colaborativa de um Dicionário Multimídia em Língua Brasileira de Sinais, em 'Anais do XVII Simpósio Brasileiro de Sistemas Multimídia e Web, Webmedia'11', Florianópolis, Brasil, pp. 244-251.
5. Lopes, L. H. L., **T. M. U. Araújo**, G. L. Souza Filho, et al. (2011), An Experience in User Interface Design for a Collaborative Web System addressed for Deaf Users, em 'Proceedings of the IADIS International Conference WWW/Internet 2011', Rio de Janeiro, Brasil, pp. 1-8

- **Dois Artigos Resumidos Publicados em Conferências Internacionais:**

1. **Araújo, T. M. U.**, A. N. Duarte & G. L. Souza Filho (2011), Accessibility as a Service: A Scalable Service for Automatic Generation of Sign Language Videos, *em* 'Proceedings of the Accessibility, Inclusion and Rehabilitation using Information Technologies, AIRtech 2011', Havana, Cuba, pp. 11-12;
2. **Araújo, T. M. U.**, T. A. Tavares & G. L. Souza Filho (2011), An architecture to generate automatic Brazilian sign language legends into Digital Television Systems, *em* 'Proceedings of the EuroITV 2009, Leuven, Bélgica, pp. 1-2.

• **Duas Patentes Depositadas:**

1. **Araújo, T. M. U.**, T. A. Tavares & G. L. Souza Filho (2009), Sistema para Automatização da Geração de Legendas em LIBRAS em Programas de TV Digital Interativa. 2009, Brasil. Patente: Privilégio de Inovação. Número do registro: PI0904582-1, data de depósito: 11/02/2009.
2. Lemos, F. H., **T. M. U. Araújo**, G. L. Souza Filho, et al. (2009), Sistema para Geração Automática de Legendas em LIBRAS em Cinemas Digitais, 2012, Brasil. Patente: Privilégio de Inovação, Data de depósito: 23/05/2012.

• **Dois Projetos Financiados:**

1. GTAAAS - Acessibilidade como um Serviço (*Accessibility as a Service*)
Resumo: Pessoas com deficiência enfrentam sérias dificuldades para acessar informações. As tecnologias de informação e comunicação (TICs) quando são desenvolvidas dificilmente levam em conta os requisitos específicos destes usuários especiais. Neste projeto exploraremos o conceito de Acessibilidade como um Serviço, implementando e testando um serviço que torne possível o acesso a conteúdo digital para usuários especiais. A proposta é adicionar uma camada entre os sistemas legados e os usuários especiais, adaptando a apresentação dos conteúdos digitais para as necessidades destes usuários. Considerando as restrições de recurso e tempo do projeto, trataremos inicialmente de acessibilidade para usuários da RNP com deficiência auditiva, um grupo de usuários que apresenta maior dificuldade de comunicação devido à escassez de soluções técnicas adequadas para a Língua Brasileira de Sinais (LIBRAS).
Função: Coordenador adjunto
Financiador: Rede Nacional de Ensino e Pesquisa (RNP)
2. ASSISTA - Criação do Núcleo Integrado de Acessibilidade, Inclusão e Desenvolvimento de Tecnologias Assistivas da UFPB
Resumo: A proposta do ASSISTA é integrar as habilidades multidisciplinares

dos pesquisadores UFPB e fomentar o desenvolvimento de projeto de PD&I na área de Tecnologia Assistiva, tendo como foco as seguintes linhas temáticas: (1) Adequação Postural, (2) Auxílio para Cegos e Surdos (Acesso à Informação) e (3) Auxílio para a Vida Cotidiana

Financiador: Secretaria de Ciência e Tecnologia para Inclusão Social (SECIS) do Ministério da Ciência, Tecnologia e Inovação (MCTI)

7.2 Propostas de Trabalhos Futuros

Embora este trabalho tenha proposto uma abordagem abrangente de tradução automática para LIBRAS, diversos desafios de pesquisa e propostas de trabalhos futuros podem ser identificados. Uma das possíveis propostas de trabalho futuro, por exemplo, está relacionada com a integração de estratégias de reconhecimento de voz na arquitetura da solução proposta. Essa integração poderia permitir a geração dos vídeos de LIBRAS a partir do áudio, e uma investigação mais profunda relacionada ao desempenho e qualidade da tradução precisaria ser explorada.

Um esforço maior também precisa ser empregado para tornar a sinalização virtual (baseada em avatares) mais natural para os surdos. Dessa forma, outra proposta de trabalho futuro envolve a incorporação de equipamentos de captura de movimentos, como, por exemplo, Microsoft Kinect², no WikiLIBRAS, permitindo que novos sinais sejam gerados a partir de capturas de movimentos realizados por humanos.

Outras propostas de trabalhos futuros que podem ser elencadas:

- Incorporar na solução mecanismos que viabilizem a revisão das traduções por colaboradores humanos. Isso permitiria estender o papel dos colaboradores na solução e possibilitar a produção de traduções com melhor qualidade para conteúdos que não exigem tradução em tempo real (tradução sob demanda, por exemplo).
- Incorporação de uma estratégia de reordenação das seqüências de glosas baseado em um modelo de linguagem extraído automaticamente de um corpus em LIBRAS (tradução automática estatísticas).
- Incorporação do uso de classificadores e de informações semânticas na estratégia de tradução.
- Desenvolvimento de uma arquitetura escalável, distribuída e tolerante a falhas para aprofundar as discussões com relação ao conceito de "Acessibilidade como um

²www.xbox.com

Serviço".

- Desenvolvimento de um protótipo da solução proposta aplicado ao ensino.
- Desenvolvimento de estratégias de tradução automática de LIBRAS para Português. A tradução no sentido inverso do proposto nesse trabalho (que propõe tradução Português para LIBRAS) permitiria que estratégias de interação e comunicação bidirecional fossem desenvolvidas para os surdos brasileiros.
- Inclusão de vídeos com intérpretes de LIBRAS para auxiliar na navegação dos usuários no WikiLIBRAS.

Referências Bibliográficas

- ABNT (2005), 'Abnt nbr 15290 acessibilidade em comunicação na televisão'.
- ABNT (2007), 'Abnt nbr 15606-1 televisão digital terrestre - codificação de dados e especificações de transmissão para radiodifusão digital - parte 1: codificação de dados'.
- Acesso Brasil (2012), 'Dicionário da língua brasileira de sinais'.
URL: <http://www.acessobrasil.org.br/libras/>
- Adobe (2012a), 'Adobe flash player 11'.
URL: <http://www.adobe.com/br/products/flashplayer.html>
- Adobe (2012b), 'Cinemática inversa'. Adobe Flash Professional.
URL: http://help.adobe.com/pt_BR/flash/cs/using/WS58BD3A02-DA25-488f-B534-AE5463A24833.html
- Ahn, L. V., B. Maurer, C. McMillen, D. Abraham & M. Blum (2008), 'Recaptcha: human-based character recognition via web security measures', *Science* **321**(5895), 1465–1468.
- Ahn, L. V. & L. Dabbish (2004), Labeling images with a computer games, *em* 'Proceedings of the SIGCHI conference on Human factors in computing systems - CHI'04', Vienna, Áustria, pp. 319–326.
- Ahn, L. V. & L. Dabbish (2008), 'Designing games with a purpose', *Communications of the ACM* **54**(8), 58–67.
- Ahn, L. V., R. Liu & M. Blum (2006), Peekaboom: a game for locating objects in images, *em* 'Proceedings of the SIGCHI conference on Human factors in computing systems - CHI'04', Montreal, Canadá, pp. 55–64.
- Anuja, K., S. Suryapriya & S. M. Idicula (2009), Design and development of a frame based mt system for english-to-isl, *em* 'Proceedings of the World Congress on Nature & Biologically Inspired Computing- NaBIC 2009', Coimbatore, Índia, pp. 1382–1387.

- Araújo, T. M. U., F. L. S. Ferreira, D. A. N. S. Silva, E. L. Falcão, L. Dantas, L. A. Domingues, L. H. Lopes, Y. Sato, H. R. Lima, A. N. Duarte & G. L. Souza Filho (2012), 'Accessibility as a service: Augmenting multimedia content with sign language video tracks', *Journal of Research and Practice in Information Technology* (Aceito para publicação).
- Batista, L. V. & M. M. Meira (2004), 'Texture classification using the lempel-ziv-welch algorithm', *Lecture Notes in Computer Science* **3171**, 444–453.
- Biolchini, J., P. G. Milan, A. C. C. Natali & G. H. Travassos (2005), Systematic review in software engineering, Relatório técnico, System Engineering and Computer Science Department COPPE/UFRJ.
- Blakowski, G. & R. Steinmetz (1996), 'A media synchronization survey: reference model, specification and case studies', *IEEE Journal on Selected Areas in Communications* **14**(1), 5–35.
- Bratko, A., B. Filipic, G. V. Cormack, T. R. Lynam & B. Zupan (2006), 'Spam filtering using statistical data compression models', *Journal of Machine Learning Research* **7**(12), 2673–2698.
- Brito, L. F. (1995), *Por uma gramática de língua de sinais*, Editora Tempo Brasileiro, Rio de Janeiro, Brasil.
- Brown, P. F., V. J. D. Pietra, S. A. D. Pietra & R. L. Mercer (1993), 'The mathematics of statistical machine translation: parameter estimation', *Computational Linguistics* **19**(2), 263–311.
- Buttussi, F., L. Chittaro & M. Coppo (2007), Using web3d technologies for visualization and search of signs in an international sign language dictionary, em 'Proceedings of the International Conference on 3D Web Technology', Perugia, Itália, pp. 61–70.
- Campos, M. B. & L. M. M. Girafa (2000), Signsim: uma ferramenta para auxílio à aprendizagem da língua brasileira de sinais, em 'Proceedings of V Congresso Ibero-Americano de Informática na Educação - RIBIE', Vinãdelmar, Chile, pp. 1–13.
- Chiu, Y. H., C. H. Wu, H.Y. Su & C.J. Cheng (2007), 'Joint optimization of word alignment and epenthesis generation for chinese to taiwanese sign synthesis', *IEEE Transactions on Advances in Computers Pattern Analysis and Machine Intelligence* **29**(1), 28–39.

- Cleary, J.G. & I. H. Witten (1984), 'Data compression using adaptive coding and partial string matching', *IEEE Transactions on Communications* **32**(4), 396–402.
- Coimbra, T. C. (2011), 'Computação humana aplicada ao problema do caixeiro viajante'. Monografia (Bacharelado em Ciências da Computação), Universidade Federal de Pelotas.
- Cox, S., M. Lincoln, J. Tryggvason, M. Nakisa, M. Wells, M. Tutt & S. Abbott (2002), Tessa, a system to aid communication with deaf people, em 'Proceedings of the Fifth International ACM Conference on Assistive Technologies - Assets'02', Edinburgh, Escócia, pp. 205–212.
- Dangsaart, S., K. Naruedomkul, N. Cercone & B. Sirinaovakul (2008), 'Intelligent thai text - thai sign translation for language learning', *Computers & Education* **51**, 1125–1141.
- Dasgupta, T. & A. Basu (2008), Prototype machine translation system from text-to-indian sign language, em 'Proceedings of of the 13th International Conference on Intelligent User Interfaces- IUI'08', Gran Canaria, Canada, pp. 313–316.
- D'Haro, L. F., R. San-Segundo, R. Córdoba, J. Bugerth, D. Stein & H. Ney (2008), Language model adaptation for a speech to sign language translation system using web frequencies and a map framework, em 'Proceedings of the Interspeech 2008', Brisbane, Austrália, pp. 2199–2202.
- Digital Cinema Initiatives, LLC (2008), 'Digital cinema system specification'.
- Dorr, B., P. Jordan & J. Benoit (1999), 'A survey of current paradigms in machine translation', *Advances in Computers* **49**, 1–68.
- Elliott, R., J. R. Glauert & J. R. Kennaway (2004), A framework for non-manual gestures in a synthetic signing system, em 'Proceedings of the of the Cambridge Workshop Series on Universal Access and Assistive Technology', Cambridge, Reino Unido, pp. 127–136.
- FACTory, The (2012), 'The factory'.
URL: <http://game.cyc.com>
- Felipe, T. A. (2007), *Libras em Contexto : Curso Básico*, 8ª edição, WalPrint Gráfica e Editora, Rio de Janeiro, Brasil.

- FGV/ABERT (2012), 'Pesquisa sobre tv digital no brasil'.
URL: <http://www.abert.org.br/site/images/stories/pdf/TVProgramacao.pdf>
- Fotinea, S. E., E. Efthimiou, G. Caridakis & K. Karpouzi (2008), 'A knowledge-based sign synthesis architecture', *Universal Access in the Information Society* **6**, 415–418.
- Freitas, C., P. Rocha & E. Bick (2008), Floresta sintá(c)tica: bigger, thicker and easier, em 'Proceedings of the 8th international conference on Computational Processing of the Portuguese Language, PROPOR'08', Aveiro, Portugal, pp. 216–219.
- Frigge, M., D. C. Hoaglin & B. Iglewicz (1989), 'Some implementations of the boxplot', *The American Statistician* **43**(1), 50–54.
- Fusco, E. (2004), X-libras: Um ambiente virtual para a língua brasileira de sinais, Dissertação de mestrado, Centro Universitário "Eurípedes de Marília", UNIVEM, Marília, SP.
- Gallo, B., R. San-Segundo, J. M. Lucas, R. Barra, L. F. D'Haro & F. Fernández (2009), 'Telecom+d04: Speech into sign language statistical translation system for deaf people', *IEEE Latin America Transactions* **7**(3), 400–404.
- Gibet, S., T. Leborque & P. F. Marteau (2001), 'High-level specification and animation of communicative gestures', *Journal of Visual Languages and Computing* **12**, 657–687.
- Goes, M. C. R. (1996), *Linguagem, Surdez e Educação*, Editora Autores Associados, Campinas, Brasil.
- Groves, R. M., F. J. Fowler Jr, M. P. Couper, J. M. Lepkowski, E. Singer & R. Tourangeau (2009), *Survey Methodology*, 2ª edição, Willey Blackwell, Hoboken, EUA.
- Haddon, L. & G. Paul (2001), *Technology and the Market: Demand, Users and Innovation*, ASEAT Conference Proceedings Series, Edward Elgar Publishing, Cheltenham, Reino Unido, capítulo Design in the ICT industry: the role of users., pp. 201–215.
- Hanke, T. (2004), Hamnosys - representing sign language data in language resources and language processing contexts, em 'Proceeding of Workshop on the Representation and Processing of Sign Languages - LREC'04', Lisboa, Portugal, pp. 1–6.

- Huenerfauth, M. (2008), 'Generating american sign language animation: overcoming misconceptions and technical challenges', *Universal Access in the Information Society* 6(4), 419–434.
- Huenerfauth, M., L. Zhao, E. Gu & J. Allbeck (2007), Evaluating american sign language generation through the participation of native asl signers, *em* 'Proceedings of the 9th international ACM SIGACCESS conference on Computers and accessibility', Tempe, EUA, pp. 211–218.
- IBGE (2000), Censo demográfico 2000: Características gerais da população, Relatório técnico, Instituto Brasileiro de Geografia e Estatística.
URL: http://www.ibge.gov.br/home/estatistica/populacao/censo2000/populacao/censo2000_populacao.pdf
- IBGE (2010), Censo demográfico 2010: Características gerais da população, religião e pessoas com deficiência, Relatório técnico, Instituto Brasileiro de Geografia e Estatística.
URL: ftp://ftp.ibge.gov.br/Censos/Censo_Demografico_2010/Caracteristicas_Gerais_Religiao_Deficiencia/tab1_3.pdf
- IEEE (2012), 'Overview articles'. IEEE Signal Processing Society.
URL: <http://http://www.signalprocessingsociety.org/publications/overview-articles>
- IGDA (2012), '2008-2009 casual games white paper'.
URL: http://www.igda.org/sites/default/files/IGDA_Casual_Games_White_Paper_2008.pdf
- ISO/IEC (1996), 'So/iec13818-1 tr information technology - generic coding of moving pictures and associated information: part 1: systems'.
- ISO/IEC (1998), 'So/iec13818-6 tr information technology - generic coding of moving pictures and associated information: part 6: extension for digital storage media command and control'.
- Jain, R. (1991), *The Art of Computer Systems Performance Analysis: Techniques for Experimental Design, Measurement, Simulation, and Modeling*, Willey, Nova York, EUA.

- Kim, S. W., Z. H. Li & Y. Aoki (2004), On intelligent avatar communication using korean, chinese and japanese sign-languages: an overview, *em* 'Proceeding of 8th International Conference on Control, Automation, Robotics and Vision, ICARCV 2004', Kunming, China, pp. 747–752.
- Kipp, M., Q. Nguyen, A. Heloir & S. Matthes (2012), Assessing the deaf user perspective on sign language avatars, *em* 'Proceeding of 13th International ACM SIGACCESS Conference on Computers and Accessibility', Dundee, Escócia, pp. 1–8.
- Kit, C., H. Pan & J. J. Webster (2002), *Translation and Information Technology*, Chinese University of HK Press, Hong Kong, China, capítulo Example-based machine translation: A new paradigm, pp. 57–78.
- Kitchenham, B. & S. Charters (2005), Guidelines for performing systematic literature reviews in software engineering, Relatório técnico, Keele University and University of Durham.
- Knight, K. (1997), 'Automating knowledge acquisition for machine translation', *AI Magazine* **18**(4), 81–96.
- Koehn, P., F. J. Och & D. Marcu (2003), Statistical phrase-based translation, *em* 'Proceeding of the Human Language Technology Conference 2003, HLT-NAACL'03', Edmonton, Canadá, pp. 127–133.
- Kumar, S., Y. Deng & W. Byrne (2004), 'A weighted finite state transducer translation template model for statistical machine translation', *Natural Language Engineering* **1**(1), 1–41.
- Lee, D. G., D. I. Fels & J. P. Udo (2007), 'Emotive captioning', *Computers in Entertainment* **5**(2), 3–15.
- Lee, S., V. Henderson, H. Hamilton, T. Starner, H. Brashear & S. Hamilton (2005), A gesture based american sign language game for deaf children, *em* 'Proceeding of the Conference on Human Factors in Computing Systems - CHI'2005', Portland, EUA, pp. 1589–1592.
- Lopez-Ludena, V., R. San-Segundo, R. Martín, D. Sánchez & A. Garcia (2011), 'Evaluating a speech communication system for deaf people', *IEEE Latin America Transactions* **9**(4), 565–570.

- Macedo, D.R. (1999), Sign dic: Um ambiente multimídia para a criação e consulta de dicionários bilíngües de línguas de sinais e línguas orais, Dissertação de mestrado, Pontifícia Universidade Católica do Rio Grande do Sul, PUC-RS, Porto Alegre, RS.
- Mafra, S. N. & G. H. Travassos (2006), Estudos primários e secundários apoiando a busca por evidências em engenharia de software, Relatório técnico, Programa de Engenharia de Sistemas e Computação - COPPE/UFRJ.
URL: <http://www.cos.ufrj.br/uploadfiles/1149103120.pdf>
- Mahoui, M., W. J. Teahan, W. J. T. Sekhar & S. Chilukuri (2008), Identification of gene function using prediction by partial matching (ppm) language models, *em* 'Proceedings of the 17th ACM conference on Information and knowledge management', Napa Valley, EUA, pp. 779–786.
- Medeiros, T. F. L., A. B. Cavalcanti, E. V. C. Lima Borges, I. L. P. Andrezza, B. E. S. Cavalcante & L. V. Batista (2011), Heart arrhythmia classification using the ppm algorithm, *em* 'Proceedings of the Biosignals and Biorobotics Conference', Vitória, Brasil, pp. 1–5.
- Moffat, A. (1990), 'Implementing the ppm data compression scheme', *IEEE Transactions on Communication* **38**(11), 1917–1921.
- Morrissey, S. (2008), Data-driven machine translation for sign languages, Tese de doutorado, Dublin City University, Dublin, Irlanda.
- Murakami, M., T. Kuroda, Y. Manabe & K. Chihara (2002), Generation of modifier representation in sign animation, *em* 'Proceedings of the Fourth International Conference Series on Disability, Virtual Reality and Associated Technologies', Veszprém, Hungria, pp. 27–32.
- Nagao, M. (1984), *Artificial and Human Intelligence*, Elsevier Science Publishers, Amsterdam, Holanda, capítulo A framework of a mechanical translation between Japanese and English by analogy principle, pp. 173–180.
- Othman, A. & M. Jemni (2011), 'Statistical sign language machine translation: from english written text to american sign language gloss', *International Journal of Computer Science Issues* **8**(5), 65–73.
- Papineni, K., S. Roukos, T. Ward & W. Zhu (2001), Bleu: a method for automatic evaluation of machine translation, *em* 'Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics', Philadelphia, EUA, pp. 311–318.

- Quadros, R. M. (1997), *Educação de Surdos - Aquisição da Linguagem*, Editora Artes Médicas, Porto Alegre, Brasil.
- Quinn, A. J. & B. B. Bederson (2011), Human computation: a survey and taxonomy of a growing field, em 'Proceedings of the 2011 annual conference on Human factors in computing systems - CHI'11', Vancouver, Canada, pp. 1403–1412.
- Russell, S. & P. Norvig (2004), *Inteligência Artificial*, 2ª edição, Editora Campus, Rio de Janeiro, Brasil.
- Sagawa, H., M. Ohki, E. Oohira, H. Ikeda & H. Fujisawa (1996), 'Pattern recognition and synthesis for a sign language translation system', *Journal of Visual Languages & Computing* 7(1), 109–127.
- San-segundo et al., R. (2008a), 'Proposing a speech to gesture translation architecture for spanish deaf people', *Journal of Visual Languages and Computing* 19(5), 523–538.
- San-segundo et al., R. (2008b), 'Speech to sign language translation system for spanish', *Speech Communication* 50(11), 1009–1020.
- San-Segundo, R., A. Pérez, D. Ortiz, L.F. D'Haro, M.I. Torres & F. Casacuberta (2007), Evaluation of alternatives on speech to sign language translation, em 'Proceedings of the Interspeech 2007', Antwerp, Bélgica, pp. 2529–2532.
- San-segundo, R., J. M. Montero, R. Córdoba, V. Sama, F. Fernández, L. F. D'Haro, V. López-Ludeña, D. Sánchez & A. Garcia (2011), 'Design, development and field evaluation of a spanish into sign language translation system', *Pattern Analysis and Applications* 15(2), 203–224.
- San-Segundo, R., R. Barra, L.F. D'Haro, J. M. Montero, R. Córdoba & J. Ferreiros (2006), A spanish speech to sign language translation system for assisting deaf-mute people, em 'Proceedings of the Interspeech 2006', Pittsburgh, EUA, pp. 1399–1402.
- Santos, G. S., M. S. Silveira & S. M. Aluisio (2009), Produção de textos paralelos em língua portuguesa e uma interlíngua em libras, em 'Proceedings of XXXVI Seminário Integrado de Software e Hardware - SEMISH'09', Bento Gonçalves, Brasil, pp. 371–385.
- Sato, S. (1993), Example-based translation of technical terms, em 'Proceedings of Fifth International Conferente on Theoretical and Methodological Issues in Machine Translation', Kyoto, Japão, pp. 58–68.

- Soares, L. F. G. & S. D. J. Barbosa (2005), *Programando em NCL 3.0: Desenvolvimento de aplicações para o middleware Ginga*, Editora Campus, Rio de Janeiro, Brasil.
- Souza Filho, G. L., L. E. C. Leite & C. E. C. F. Batista (2007), 'Ginga-j: The procedural middleware for the brazilian digital tv system', *Journal of the Brazilian Computer Society* **12**(4), 47–56.
- Starner, T., A. Pentland & J. Weaver (1998), 'Real-time american sign language recognition using desk and wearable computer based video', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **20**(12), 1371–1375.
- Stokoe, W. C. (1980), 'Sign language structure', *Annual Review of Anthropology* **9**, 365–390.
- Stumpf, M. R. (2000), Língua de sinais: escrita dos surdos na internet, em 'Proceedings of V Congresso Ibero-Americano de Informática na Educação', Viñadelmar, Chile, pp. 1–8.
- Su, H. Y. & C. H. Wu (2009), 'Improving structural statistical machine translation for sign language with small corpus using thematic role templates as translation memory', *IEEE Transactions on Machine Translation, Audio, Speech, and Language Processing* **17**(7), 1305–1315.
- Surowiecki, J. (2005), *Wisdom of Crowds*, Anchor, Nova York, EUA.
- Sutton, V. (1995), *Lessons in Sign Writing, Textbook and Workbook*, 2ª edição, Deaf Action Committee for SignWriting, La Jolla, EUA.
- van Zijl, L. & G. Olivrin (2008), South african sign language assistive translation, em 'Proceedings of the IASTED International Conference on Telehealth/Assistive Technologies', Baltimore, EUA, pp. 3–7.
- Veale, T., A. Conway & B. Collins (1998), 'The challenges of cross-modal translation: English to sign language translation in the zardo system', *Machine Translation* **13**(1), 81–106.
- Wauters, L. N. (2005), Reading comprehension in deaf children: The impact of the mode of acquisition of word meanings, Tese de doutorado, Radboud University, Nijmegen, Holanda.

- WHO (2012), 'Deafness and hearing impairment'. World Health Organization.
URL: <http://www.who.int/mediacentre/factsheets/fs300/en/>
- Wohlin, C., P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell & A. Wesslén (2000), *Experimentation in Software Engineering: An Introduction*, Kluwer Academic Publisher, Norwell, EUA.
- Woodward, J.C. (1973), 'Some observation on sociolinguistic variation and american sign language', *Kansas Journal of Sociology* **9**(2), 191–200.
- Wu, C. H., H. Y. Su, Y. H. Chiu & C. H. Lin (2007), 'Transfer-based statistical translation of taiwanese sign language using pcfg', *ACM Transactions on Asian Language Information Processing* **16**(1), 1–18.
- Yuen, M. C., L. J. Chen & I. King (2009), A survey of human computation systems, *em* 'Proceedings of the International Conference on Computational Science and Engineering -CSE '09', Vancouver, Canadá, pp. 723–728.
- Zhang, Q., L. Cheng & R. Boutaba (2010), 'Cloud computing: state-of-the-art and research challenges', *Journal of Internet Services and Applications* **1**, 7–18.
- Zhao, L., K. Kipper, W. Schuler, C. Vogler, N. Badler & M. Palmer (2000), Machine translation system from english to american sign language, *em* 'Proceedings of the 4th Conference of the Association for Machine Translation in the Americas', Cuernavaca, Mexico, pp. 54–67.

Apêndice A

Protocolo da Revisão Sistemática

Conforme mencionado no início do Capítulo 3, por serem baseadas numa abordagem sistemática, as revisões sistemáticas, em geral, se iniciam definindo um **protocolo de revisão sistemática**. Esse protocolo representa um plano para conduzir a revisão sistemática e descreve qual é a questão de pesquisa abordada, quais são os métodos e estratégias de buscas utilizados para realizar a revisão, dentre outros.

O primeiro passo no processo de definição desse protocolo é identificar/definir quais são as "questões de pesquisa". Considerando que o tema de pesquisa dessa revisão são os sistemas de tradução automática para línguas de sinais, as questões de pesquisa definidas nesse estudo foram as seguintes:

- QP 1: Como estão distribuídas as publicações relacionadas a tradução automática para língua de sinais?
- QP 2: Podemos distribuir essas publicações segundo uma taxonomia?
- QP 3: Quais são as principais contribuições dos sistemas de tradução automática de línguas orais para línguas de sinais?
- QP 4: Quais as formas de avaliação comumente utilizadas para analisar a qualidade de técnicas de tradução automática para línguas de sinais? Quais os principais resultados obtidos?

A pesquisa sobre trabalhos relacionados¹ foi realizada através de um processo de busca automática sobre os principais portais (bibliotecas digitais) de artigos científicos. Nesse estudo foram selecionados os seguintes portais de artigos científicos:

- IEEEXplore (<http://ieeexplore.ieee.org>)
- ACM Digital Library (<http://dl.acm.org>)
- ScienceDirect (<http://www.sciencedirect.com/>)
- Scopus (<http://www.scopus.com/>)

¹Nesse estudo, foram considerados apenas documentos escritos em inglês.

- ISI Web of Knowledge (<http://isiwebofknowledge.com/>)

Uma característica importante desses portais de artigos é que eles permitem realizar buscas considerando exclusivamente o título, resumo e palavras-chaves dos artigos. Com isso é possível evitar fontes em que termos que compõem a "string de busca" aparecem soltos no corpo do artigo, o que poderia resultar na seleção de trabalhos que não estão diretamente relacionados com o tema de pesquisa.

Para aplicar o processo de busca automática, além das fontes de pesquisa (portais de artigos), também é necessário definir uma "string de busca". Para formar essa string, é necessário antes, no entanto, definir os "termos de pesquisa", isto é, as palavras chaves, sinônimos e termos relacionados que irão compor essa string de busca. Esses termos, apresentados na Tabela A.1, juntamente com suas variantes foram então combinados, formando a seguinte "string de busca":

- ("sign language" OR "visual language" OR LIBRAS OR ASL) AND (translat* OR pars* OR "language processing")

Tabela A.1: Termos de busca

Palavras-chaves	Sinônimos e termos relacionados
sign language	visual language, ASL, LIBRAS
machine translation	translation, parsing, language processing

Após a aplicação do processo de busca, os documentos recuperados são filtrados com base nos seguintes critérios de inclusão (I) e exclusão (E):

- E1: Excluir todos os documentos que não são artigos científicos;
- E2: Excluir os trabalhos repetidos. Caso existam trabalhos repetidos, incluir apenas o trabalho mais recente;
- E3: Excluir trabalhos em que o acesso ao artigo completo não esteja disponível, isto é, apenas o resumo (*abstract*) está disponível;
- I1: Incluir todos os artigos completos relacionados ao contexto da revisão (através da leitura do título e do resumo e, em caso de informações insuficientes, da leitura completa do artigo);

Após esse processo de filtragem, os artigos selecionados são classificados com os seguintes critérios:

1. **Estratégia de tradução:** Identifica qual o método de tradução automática utilizado pelo autor do trabalho. Considerando esse critério, os artigos podem ser classificados da seguinte forma:
 - (a) Baseada em Regras: utiliza apenas estratégias baseadas em Tradução Automática Baseada em Regras;
 - (b) Estatística: utiliza apenas Tradução Automática Estatística;
 - (c) Baseada em Exemplos: utiliza apenas Tradução Automática Baseada em Exemplos;
 - (d) Híbridos: utiliza mais de um destes paradigmas;
 - (e) Sem classificação: a estratégia de tradução não é identificada ou definida no artigo;

2. **Tipo de Entrada:** Identifica o tipo de entrada utilizado pelo sistema de tradução. Nesse caso, os tipos de entradas mais comumente utilizadas são:
 - (a) Voz: aborda a tradução a partir de voz ou fala;
 - (b) Texto: aborda tradução a partir de mídias textuais;
 - (c) Sem classificação: o tipo de entrada não é identificado ou definida no artigo;

3. **Tipo de Saída:** Identifica o tipo de saída gerado pelo sistema de tradução. Os tipos de saída mais comumente utilizados são:
 - (a) Animação: gera uma saída baseada em agentes animados virtuais (avatars);
 - (b) Vídeos com intérpretes: gera uma saída a partir vídeos pré-gravados por intérpretes de línguas de sinais;
 - (c) Glosa: gera uma saída em glosa (representação textual em língua de sinais);
 - (d) Representação Escrita em Língua de Sinais: gera uma saída baseada nas representações escritas de Stokoe, HamNoSys ou SignWriting (ver Seção 2.1.3);
 - (e) Sem classificação: quando o tipo de saída não é identificado ou definida no artigo;

4. **Domínio da Tradução:** Identifica se o sistema de tradução foi projetado para traduzir qualquer tipo de texto (notícias, esportes, política, clima, filmes, educação, etc.) ou se ele foi desenvolvido para traduzir textos em um domínio de aplicação específico. Nesse contexto, com base nesse critério, o artigo pode ser classificado da seguinte forma:
 - (a) Domínio geral: envolve traduções para qualquer tipo de texto;
 - (b) Específico: envolve traduções de conteúdos para um domínio específico;

- (c) Sem classificação: o domínio da tradução não é identificado ou definido no artigo;
5. **Método de avaliação:** Identifica a estratégia de avaliação utilizado para avaliar o desempenho, viabilidade e qualidade das traduções/conteúdos gerados pelos trabalhos. Nesse caso, a avaliação pode ser:
- (a) Automática: utiliza apenas medidas objetivas obtidas a partir de testes computacionais;
 - (b) Com usuários: realiza apenas avaliação subjetiva através de testes com usuários;
 - (c) Híbrida: envolve tanto avaliação automática quanto avaliação com usuários;
 - (d) Sem classificação: o tipo de avaliação não é definida no artigo;
6. **Presença de corpus:** Identifica se os trabalhos utilizam um corpus textual e qual a característica desse corpus. Nesse caso, os artigos podem ser classificados da seguinte forma:
- (a) Sem corpus: não utiliza nenhum corpus textual;
 - (b) Unilingue (língua oral): utiliza um corpus textual apenas na língua oral (isto é, não utiliza um corpus na língua de sinais);
 - (c) Unilingue (língua de sinais): utiliza um corpus textual apenas na língua de sinais (isto é, não utiliza um corpus em língua oral);
 - (d) Bilingüe (língua oral): utiliza um corpus textual bilingüe (isto é, nas duas línguas);
 - (e) Sem avaliação: o trabalho não apresenta qualquer tipo de avaliação sobre a solução;
7. **Tipo de artigo:** Identifica o objetivo do artigo. Nesse caso, o artigo pode ser classificado da seguinte forma:
- (a) Pesquisa ou *Survey*: artigos apresentados como *surveys*. De acordo com Groves et al. (2009), *surveys* são métodos sistemáticos de coleta de informações a partir de uma amostra de indivíduos, com o objetivo de descrever os atributos de uma população maior;
 - (b) *Overview*: artigos apresentados como *overview*. De acordo com o IEEE (2012), *overview* são artigos que pretendem fazer uma análise técnica profunda sobre um determinado tema de pesquisa;
 - (c) Experimental: artigos que reportam experimentos científicos com experimentos, objetivos, metodologia e resultados;

- (d) Sistema, descrição, modelo, etc: artigos que descrevem estratégias de tradução automática para línguas de sinais;

Apêndice B

Artigos da Revisão Sistemática

1. Abdallah, M (1998), A neuro-hierarchical multilayer network in the translation of the American sign language, *em* 'Proceedings of the IEEE Southeastcon', Orlando, EUA , pp. 224-227
2. Afzal, H., S. Nafees, S. Nasir & A. Tariq (2009), iPSL: enabling rehabilitation of deaf community in Pakistan, *em* 'Proceedings of the 3rd International Convention on Rehabilitation Engineering & Assistive Technology, i-CREATE'09', Singapura, pp. 1-4
3. Allen, J. M. & R. A. Foulds (2004), An approach to animating sign language: a spoken English to sign English translator system, *em* 'Proceedings of the 30th Annual Northeast IEEE Bioengineering Conference', Springfield, EUA, pp. 43-44
4. Anuja, K., S. Suryapriya & S. M. Idicula (2009), Design and development of a frame based MT system for english-to-isl, *em* 'Proceedings of the World Congress on Nature & Biologically Inspired Computing- NaBIC 2009', Coimbatore, Índia, pp. 1382-1387
5. Bangham, J. A., S. J. Cox, R. Elliot, J R. W. Glauert, I. Marshal, S. Rankov & M. Wells (2000), Virtual signing: capture, animation, storage and transmission-an overview of the ViSiCAST project, *em* 'Proceedings of the IEE Seminar on Speech and Language Processing for Disabled and Elderly People', Londres, Reino Unido, pp. 1-7
6. Chiu, Y. H., C. H. Wu, H.Y. Su & C.J. Cheng (2007), 'Joint optimization of word alignment and epenthesis generation for chinese to taiwanese sign synthesis', *IEEE Transactions on Advances in Computers Pattern Analysis and Machine Intelligence* **29**(1), 28-39.
7. Chung, J.W., H. J. Lee & J. C. Park (2011), Improving accessibility to web documents for the aurally challenged with sign language animation, *em* 'Proceedings of the International Conference on Web Intelligence, Mining and Semantics, WIMS'11', Sogndal, Noruega, pp. 1-8.
8. Cox, S., M. Lincoln, M. Tryggvason, M. Nakisa, M. Wells, M. Tutt & S. Abbott (2002), Tessa, a system to aid communication with deaf people, *em* 'Proceedings of the 5th International ACM Conference on Assistive Technologies, Assets'02', Edinburgh, Escócia, pp. 205-212.

9. Cox, S., M. Lincoln, M. Nakisa, M. Wells, M. Tutt & S. Abbott (2003), 'The development and evaluation of a speech to sign translation system to assist transactions', *International Journal of Human Computer Interaction* **16**(2), 141-161.
10. Dangsaart, S., K. Naruedomkul, N. Cercone & B. Sirinaovakul (2008), 'Intelligent thai text - thai sign translation for language learning', *Computers & Education* **51**, 1125-1141.
11. Dasgupta, T. & A. Basu (2008), Prototype machine translation system from text-to-indian sign language, em 'Proceedings of of the 13th International Conference on Intelligent User Interfaces-IUI'08', Gran Canaria, Canada, pp. 313-316.
12. D'Haro, L. F., R. San-Segundo, R. Córdoba, J. Bugerth, D. Stein & H. Ney (2008), Language model adaptation for a speech to sign language translation system using web frequencies and a map framework, em 'Proceedings of the Interspeech 2008', Brisbane, Austrália, pp. 2199-2202.
13. Elliot, R., J. R. W. Glauert, J. R. Kennaway & I. Marshall (2000), The development of language processing support for the ViSiCAST project, em 'Proceedings of the 4th International ACM conference on Assistive technologies, Assets'00', Arlington, EUA, pp. 101-108.
14. Elliot, R., J. R. W. Glauert, J. R. Kennaway, I. Marshall & E. Safar (2008), 'Linguistic modelling and language-processing technologies for Avatar-based sign language presentation', *Universal Access in the Information Society* **6**(4), 375-391.
15. Fotinea, S. E., E. Efthimiou, G. Caridakis & K. Karpouzi (2008), 'A knowledge-based sign synthesis architecture', *Universal Access in the Information Society* **6**, 415-418.
16. Gallo, B., R. San-Segundo, J. M. Lucas, R. Barra, L. F. D'Haro & F. Fernández (2009), 'Telecom+d04: Speech into sign language statistical translation system for deaf people', *IEEE Latin America Transactions* **7**(3), 400-404.
17. Holden, E. J. & G. G. Roy (1992), Learning tool for signed English, em 'Proceedings of the ACM/SIGAPP Symposium on Applied computing: technological challenges of the 1990's', Kansas City, EUA, pp. 444-449.
18. Hou, J. & Y. Aoki (2004a), 'A real-time interactive nonverbal communication system through semantic feature extraction as an interlingua', *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans* **34**(1), 148-154.
19. Hou, J. & Y. Aoki (2004b), A visual intercultural nonverbal communication system for the hearing impaired, em 'Proceedings of the 6th International Conference on Signal Processing', Beijing, China, pp. 985-989.

20. Huenerfauth, M. (2004), A multi-path architecture for machine translation of English text into American Sign Language animation, *em* 'Proceedings of the Student Research Workshop at HLT-NAACL, HLT-SRWS'04', Boston, EUA, pp. 25-30.
21. Huenerfauth, M. (2005a), American sign language generation: multimodal NLG with multiple linguistic channels, *em* 'Proceedings of the ACL Student Research Workshop', Ann Arbor, EUA, pp. 37-42.
22. Huenerfauth, M. (2005b), Representing coordination and non-coordination in an american sign language animation, *em* 'Proceedings of the 7th international ACM SIGACCESS conference on Computers and accessibility, Assets'05', Baltimore, EUA, pp. 44-51.
23. Huenerfauth, M. (2008), 'Generating american sign language animation: overcoming misconceptions and technical challenges', *Universal Access in the Information Society* **6**(4), 419-434.
24. Huenerfauth, M., L. Zhao, E. Gu & J. Allbeck (2007), Evaluating american sign language generation through the participation of native asl signers, *em* 'Proceedings of the 9th international ACM SIGACCESS conference on Computers and accessibility', Tempe, EUA, pp. 211-218.
25. Kamata, K., T. Yoshida, W. Watanabe & Y. Usui (1989), An approach to Japanese-sign language translation system, *em* 'Proceedings of the IEEE International Conference on Systems, Man and Cybernetics', Cambridge, EUA, pp. 1089-1090.
26. Karpouzis, K., G. Caridakis, S. E. Fotinea & E. Efthimiou (2007), 'Educational resources and implementation of a Greek sign language synthesis architecture', *Computers & Education* **49**(1), 54-74.
27. Kato, N., H. Kaneko, S. Inou, T. Shimizu, & N. Hiruma (2011), 'Machine translation to sign language with CG-animation', *ABU Technical Review* **245**, 3-6.
28. Kennaway, J., R., J. R. W. Glauert & I. Zwitterlood (2007), 'Providing signed content on the Internet by synthesized animation', *ACM Transactions on Computer-Human Interaction* **14**(3), 1-29.
29. Kim, S.W., Z. H. Li & Y. Aoki (2004), On intelligent avatar communication using korean, chinese and japanese sign-languages: an overview, *em* 'Proceeding of 8th International Conference on Control, Automation, Robotics and Vision, ICARCV 2004', Kunming, China, pp. 747-752.
30. Lee, J. & T. L. Kunii (1992), Visual translation: from native language to sign language, *em* 'Proceeding of the IEEE Workshop on Visual Languages', Seattle, EUA, pp. 103-109.
31. Lesmo, L., A. Mazzei & D. P. Radicioni (2011), An ontology based architecture for translation, *em* 'Proceeding of 9th International Conference on Computational Semantics, IWCS'11', Oxford, Reino Unido, pp. 345-349.

32. Li, J. & Y. Aoki & S. W. Kim (1998), Development of a sign language communication system between Japanese and Korean, *em* 'Proceeding of the 4th International Conference on Signal Processing Proceedings, ICSP '98', Beijing, China, pp. 916-919.
33. Lin, X. & Y. Bao-zong, G. Wen, T. Xiao-fang & X. Yan (2002), Research on machine translation oriented transformation from Chinese into Chinese sign language, *em* 'Proceeding of the 6th International Conference on Signal Processing', Beijing, China, pp. 1023-1026.
34. Lincoln, M. & S. Cox (2003), A comparison of language processing techniques for a constrained speech translation system, *em* 'Proceeding of IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP'03', Hong Kong, China, pp. 612-615.
35. Lopez-Ludena, V., R. San-Segundo, R. Martín, D. Sánchez & A. Garcia (2011), 'Evaluating a speech communication system for deaf people', *IEEE Latin America Transactions* **9**(4), 565-570.
36. Losson, O. & B. Cantegrit (2002), Generation of signed sentences by an avatar from their textual description, *em* 'Proceedings of the IEEE International Conference on Systems, Man and Cybernetics', Hammamet, Tunísia, pp. 1-6.
37. Marek, H., et al. (2011), 'Automatic fingersign-to-speech translation system', *Journal on Multimodal User Interfaces* **4**(2), 61-79.
38. Marshall, I. & E. Sáfár (2003), A prototype text to British Sign Language (BSL) translation system, *em* 'Proceedings of the 41st Annual Meeting on Association for Computational Linguistics', Sapporo, Japão, pp. 113-116.
39. Murakami, M., T. Kuroda, Y. Manabe & K. Chihara (2002), Generation of modifier representation in sign animation, *em* 'Proceedings of the 4th International Conference Series on Disability, Virtual Reality and Associated Technologies', Veszprém, Hungria, pp. 27-32.
40. Patel, I. & Y. S. Rao (2010), Technologies automated speech recognition approach to finger spelling, *em* 'Proceedings of the International Conference on Computing Communication and Networking Technologies, ICCCNT'10', Zurique, Suíça, pp. 1-6.
41. Pezeshkpour, F., I. Marshall, R. Elliot & J. A. Bangham (1999), Development of a legible deaf-signing virtual human, *em* 'Proceedings of the IEEE International Conference on Multimedia Computing and Systems', Florença, Itália, pp. 333-338.
42. Sagawa, H., M. Ohki, E. Oohira, H. Ikeda & H. Fujisawa (1996), 'Pattern recognition and synthesis for a sign language translation system', *Journal of Visual Languages & Computing* **7**(1), 109-127.
43. San-segundo et al., R. (2008a), 'Proposing a speech to gesture translation architecture for spanish deaf people', *Journal of Visual Languages and Computing* **19**(5), 523-538.

44. San-segundo et al., R. (2008b), 'Speech to sign language translation system for spanish', *Speech Communication* **50**(11), 1009-1020.
45. San-Segundo, R., A. Pérez, D. Ortiz, L.F. D'Haro, M.I. Torres & F. Casacuberta (2007), Evaluation of alternatives on speech to sign language translation, *em 'Proceedings of the Interspeech 2007'*, Antwerp, Bélgica, pp. 2529-2532.
46. San-Segundo, R., R. Barra, L.F. D'Haro, J. M. Montero, R. Córdoba & J. Ferreiros (2006), A spanish speech to sign language translation system for assisting deaf-mute people, *em 'Proceedings of the Interspeech 2006'*, Pittsburgh, EUA, pp. 1399-1402.
47. Sarkar, B., K. Datta, C. D. Datta, D. Sarkar, S. J. Dutta, A. Paul & J. U. Molla (2009), A translator for bangla text to sign language, *em 'Proceedings of the IEEE Annual India Conference, INDICON'*, Ahmedabad, Índia, pp. 1-4.
48. Su, H. Y. & C. H. Wu (2009), 'Improving structural statistical machine translation for sign language with small corpus using thematic role templates as translation memory', *IEEE Transactions on Machine Translation, Audio, Speech, and Language Processing* **17**(7), 1305-1315.
49. va Zijl, L. & A. Combrink (2006), The South African sign language machine translation project: issues on non-manual sign generation, *em 'Proceedings of the 2006 annual research conference of the South African institute of computer scientists and information technologists on IT research in developing countries, SAICSIT '06'*, Gordon's Bay, África do Sul, pp. 127-134.
50. va Zijl, L. & D. Barker (2003), South African sign language machine translation system, *em 'Proceedings of the 2nd international conference on Computer graphics, virtual Reality, visualisation and interaction in Africa, AFRIGRAPH'03'*, Cape Town, África do Sul, pp. 49-52.
51. va Zijl, L. & G. Olivrin (2008), South african sign language assistive translation, *em 'Proceedings of the IASTED International Conference on Telehealth/Assistive Technologies'*, Baltimore, EUA, pp. 3-7.
52. Veale, T. & A. Conway (1994), 'Cross modal comprehension in ZARDOZ an English to sign-language translation system', *em 'Proceedings of the 7th International Workshop on Natural Language Generation, INLG'94'*, Kennebunkport, EUA, pp. 249-252.
53. Veale, T., A. Conway & B. Collins (1998), 'The challenges of cross-modal translation: English to sign language translation in the zardoz system', *Machine Translation* **13**(1), 81-106.
54. Wu, C. H., H. Y. Su, Y. H. Chiu & C. H. Lin (2007), 'Transfer-based statistical translation of taiwanese sign language using pcfgr', *ACM Transactions on Asian Language Information Processing* **16**(1), 1-18.

55. Xu, L. & W. Gao (2000), 'Study on translating Chinese into Chinese sign language', *Journal of Computer Science and Technology* **15**(5), 485-490.
56. Zhao, L., K. Kipper, W. Schuler, C. Vogler, N. Badler & M. Palmer (2000), A machine translation system from english to american sign language, *em* 'Proceedings of the 4th Conference of the Association for Machine Translation in the Americas', Cuernavaca, Mexico, pp. 54-67.

Apêndice C

Parâmetros da LDS

Tabela C.1: Tipos de trajetórias de movimento

Tipos
retilíneo
circular
semi-circular
pontual
espiral
dobramento-pulso
torção-pulso
contato

Tabela C.2: Ponto de Articulação

Orientação da palma	Sentido da palma	Sentido dos dedos
paralela-ao-corpo	para-trás	para-cima
paralela-ao-corpo	para-trás	para-baixo
paralelo-ao-corpo	para-dentro	para-cima
paralelo-ao-corpo	para-dentro	para-baixo
paralelo-ao-corpo	para-frente	para-cima
paralelo-ao-corpo	para-frente	para-baixo
paralelo-ao-solo	para-baixo	para-frente
paralelo-ao-solo	para-baixo	para-o-corpo
paralelo-ao-solo	para-dentro	para-frente
paralelo-ao-solo	para-cima	para-frente
paralelo-ao-solo	para-cima	para-o-corpo

Tabela C.3: Ponto de Articulação

Localização	Subdivisão
cabeça	olho-direito
cabeça	olho-esquerdo
cabeça	orelha-direita
cabeça	orelha-esquerda
cabeça	queixo
cabeça	boca
cabeça	testa
corpo	busto-direito
corpo	busto-esquerdo
corpo	cintura-direita
corpo	cintura-esquerda
corpo	umbigo
corpo	ombro-direito
corpo	ombro-esquerdo
espaço-neutro	*

Tabela C.4: Expressões faciais

Tipos de Expressões
assustado
bico
bochechas-infladas
bochechas-contraídas
com-raiva
desconfiado
expressão-neutra
lábios-contraídos
língua-para-baixo
língua-para-fora
olhos-arregalados
olhos-para-esquerda
olhos-para-direita
sobrancelhas-levantadas
triste

Apêndice D

Planejamento dos Experimentos

Nesse Apêndice serão apresentados a definição e o planejamento do experimento conduzido para avaliar a solução. Conforme mencionado no Capítulo 6, a fase de definição e planejamento do experimento envolvem a definição do propósito do experimento, assim como o seu contexto, a formulação das hipóteses, definição das variáveis, sujeitos, objetos e instrumentos do experimento, dentre outros.

D.1 Definição do Experimento

Conforme mencionado no Capítulo 6, a motivação deste experimento é avaliar a solução proposta em três partes definidas de acordo com as hipóteses gerais formuladas no Capítulo 1: (1) nível de compreensão dos conteúdos gerados pela solução proposta por usuários surdos, (2) desempenho de tradução da solução proposta e (3) produtividade dos usuários na geração das construções lingüísticas da solução proposta. Nas próximas subseções, os elementos de cada parte desse experimento são definidos.

D.1.1 Avaliação do Nível de Compreensão dos Conteúdos

Com relação ao nível de compreensão e à qualidade dos conteúdos gerados pela solução proposta, os elementos que caracterizam essa parte do experimento foram definidos da seguinte forma:

- **Objeto de estudo:** O objeto de estudo são os conteúdos acessíveis gerados pela solução proposta.
- **Propósito:** O propósito dessa parte do experimento é avaliar os conteúdos gerados pela solução proposta com relação a capacidade de compreensão pelos usuários surdos. Para isso, o nível de compreensão dos conteúdos acessíveis gerados pela

solução proposta foi avaliada e confrontada com o nível de compreensão de conteúdos com legendas.

- **Foco de qualidade:** Os principais efeitos estudados pelo experimento são o nível de compreensão dos conteúdos, a qualidade de tradução e apresentação e a naturalidade dos conteúdos gerados.
- **Perspectiva:** A partir do ponto de vista dos usuários surdos brasileiros.
- **Contexto:** Essa parte do experimento foi conduzida em um ambiente controlado na Fundação Centro Integrado de Apoio à Pessoa com Deficiência (Funad) da Paraíba. Os sujeitos do experimento foram usuários surdos brasileiros e os objetos foram conteúdos multimídia, de diferentes gêneros, gerados em duas versões diferentes: (1) com trilhas de LIBRAS geradas pela solução proposta e (2) com legendas. Os usuários foram divididos aleatoriamente em dois grupos de forma que cada grupo avaliasse apenas uma das versões acessíveis de cada um dos conteúdos.

Dessa forma, essa parte do experimento pode ser definida resumidamente da seguinte forma:

- Analise conteúdos multimídia acessíveis gerados pela solução proposta
- com o propósito de avaliação
- com respeito ao nível de compreensão dos conteúdos gerados, qualidade de tradução, qualidade de apresentação e naturalidade.
- a partir do ponto de vista dos surdos brasileiros
- no contexto de surdos brasileiros avaliando conteúdos multimídia dentro de um ambiente controlado (Funad-PB).

D.1.2 Avaliação do Desempenho de Tradução

O experimento também abordou o desempenho de tradução da solução proposta. Nesse contexto, os elementos do experimento foram definidos da seguinte forma:

- **Objeto de estudo:** O objeto de estudo é o desempenho de tradução da solução proposta.
- **Propósito:** O propósito é avaliar o desempenho da solução proposta para gerar conteúdos acessíveis em um cenário que exige tradução dos conteúdos em tempo real.
- **Foco de qualidade:** O principal efeito estudado é o tempo médio para traduzir diversos conteúdos multimídia usando a solução proposta nesses cenários.

- **Perspectiva:** A partir do ponto de vista dos pesquisadores.
- **Contexto:** Essa parte do experimento foi conduzida através de um conjunto de testes computacionais automáticos realizados no ambiente da TV Digital. Para isso, o LibrasTV, protótipo da solução proposta desenvolvido para TV Digital (ver Seção 5.1), foi integrado no ambiente da TV e o desempenho de tradução foi avaliado para um cenário real de tradução.

Essa parte do experimento, portanto, pode ser definida resumidamente da seguinte forma:

- Analise o desempenho de tradução da solução proposta
- com o propósito de avaliação
- com respeito ao tempo médio de tradução e geração dos conteúdos acessíveis
- a partir do ponto de vista dos pesquisadores
- no contexto de testes computacionais objetivos realizados em um cenário real de tradução na TV Digital.

D.1.3 Avaliação da Ferramenta de Computação Humana

Com relação à avaliação da ferramenta de colaboração e computação humana, o experimento foi definido da seguinte forma:

- **Objeto de estudo:** Ferramenta WikiLIBRAS
- **Propósito:** O propósito é avaliar a produtividade dos usuários surdos e intérpretes de LIBRAS na geração das construções linguísticas no WikiLIBRAS. Para isso, a eficácia (número de sinais gerados corretamente) e eficiência (tempo médio para gerar cada sinal) de surdos e intérpretes de LIBRAS na geração de sinais em LIBRAS usando a ferramenta, foi confrontada com a eficácia e a eficiência de *designers-3D* gerando os mesmos sinais manualmente em ferramentas de modelagem e animação.
- **Foco de qualidade:** Os principais efeitos estudados pelo experimento são a eficácia e a eficiência do uso da ferramenta na geração de sinais em LIBRAS.
- **Perspectiva:** A partir do ponto de vista de surdos brasileiros e intérpretes de LIBRAS.
- **Contexto:** Essa parte do experimento também foi conduzida na Funad-PB com um conjunto de usuários surdos brasileiros, intérpretes de LIBRAS e *designers-3D*. Esses usuários foram convidados a gerar um conjunto de sinais em LIBRAS e a sua eficiência e eficácia na geração desses sinais foi avaliada.

Resumidamente, essa parte do experimento pode ser definida da seguinte forma:

- Analise a ferramenta WikiLIBRAS
- com o propósito de avaliação
- com respeito a eficácia e a eficiência na geração de sinais em LIBRAS
- a partir do ponto de vista dos surdos brasileiros e intérpretes de LIBRAS
- no contexto de surdos brasileiros e intérpretes de LIBRAS gerando sinais no WikiLIBRAS e *designers-3D* gerando esses mesmos sinais manualmente em ferramentas de modelagem e animação.

D.2 Planejamento do Experimento

Nessa seção, o planejamento de cada uma das partes do experimento será descrito. Nela será apresentado o contexto do experimento, as hipóteses que foram testadas, os sujeitos, variáveis e objetos utilizados no experimento, dentre outros.

D.2.1 Avaliação do Nível de Compreensão dos Conteúdos

Seleção do Contexto

Essa parte do experimento foi conduzida na Funad-PB com um conjunto de vinte surdos brasileiros selecionados por conveniência. Esses usuários foram convidados a avaliar quatro conteúdos multimídia gerados em duas versões diferentes: (1) com legendas ou (2) com trilhas de LIBRAS geradas pela solução proposta. Os usuários foram divididos aleatoriamente em dois grupos de forma que cada usuário avaliasse apenas uma das versões de cada um dos conteúdos. A amostra de conteúdos multimídia era heterogênea com conteúdos que pertenciam a diferentes gêneros, como, por exemplo, trechos de vídeos de noticiários, de filmes, de programas infantis e de programas de variedades, e os conteúdos eram curtos para que a avaliação não fosse cansativa e tivesse influência no resultado da avaliação.

Formulação das Hipóteses

Informalmente, as hipóteses formuladas nessa parte do experimento foram as seguintes:

1. Os conteúdos acessíveis baseados na solução proposta neste trabalho tornam os conteúdos mais bem compreendidos pelos surdos do que os conteúdos com legendas.

2. Os conteúdos acessíveis baseados na solução proposta neste trabalho têm uma qualidade de tradução melhor do que os conteúdos com legendas.

A partir das definições informais, foi possível derivar uma definição formal das hipóteses e definir quais seriam as medidas necessárias para avaliá-las.

1. **Hipótese nula H_0 :** Os conteúdos acessíveis baseados na solução proposta (SP) são tão bem compreendidos pelos usuários quanto os conteúdos com legendas (Leg), onde o nível de compreensão dos conteúdos foi medido utilizando um conjunto de perguntas sobre os conteúdos avaliados.

$$\mu_{compreensao}(SP) = \mu_{compreensao}(Leg) \quad (D.1)$$

Hipótese alternativa H_1 : Os conteúdos acessíveis baseados na solução proposta (SP) são mais bem compreendidos do que os conteúdos com legendas (Leg).

$$\mu_{compreensao}(SP) > \mu_{compreensao}(Leg), \quad (D.2)$$

2. **Hipótese nula H_0 :** Considerando a adequação a gramática de LIBRAS, os conteúdos acessíveis baseados na solução proposta (SP) são tão compatíveis com a gramática de LIBRAS quanto os conteúdos com legendas, onde o nível de adequação (compatibilidade) a gramática foi avaliado pelos usuários utilizando uma escala de 1 a 6.

$$\mu_{traducao}(SP) = \mu_{traducao}(Leg) \quad (D.3)$$

Hipótese alternativa H_1 : Os conteúdos baseados na solução proposta (CASP) são mais compatíveis a gramática de LIBRAS.

$$\mu_{traducao}(SP) > \mu_{traducao}(Leg), \quad (D.4)$$

Seleção das Variáveis

A **variável independente** desse estudo é o tipo de conteúdo acessível avaliado pelo usuário. Ela é uma variável nominal que pode assumir os seguintes valores: (SP) conteúdos acessíveis com trilhas de LIBRAS gerados pela solução proposta ou (Leg) conteúdos acessíveis com legendas.

As **variáveis dependentes** são as variáveis de resposta aos diferentes valores (tratamentos) assumidos pela variável independente. Neste estudo, as variáveis dependentes

são o nível de compreensão dos conteúdos e a adequação a gramática (qualidade de tradução). Adicionalmente, a qualidade de apresentação e a naturalidade dos conteúdos gerados pela solução proposta também são avaliados.

Seleção dos Sujeitos

Os sujeitos (usuários) desse experimento são vinte usuários surdos voluntários e escolhidos por conveniência. Os usuários possuem uma diversificada faixa etária e diferentes graus de escolaridade e são alunos dos cursos de formação em LIBRAS oferecidos pela Funad-PB.

Projeto do Experimento

Segundo os padrões de projetos de experimentos apresentados por Wohlin et al. (2000), a definição, hipótese e as medidas indicam que o projeto desta parte do experimento deve ser do tipo **um fator com dois tratamentos** (do inglês, **one factor with two treatments**), onde o fator seria o tipo de conteúdo acessível e os tratamentos seriam as possíveis escolhas para esse fator: SP ou Leg.

Os princípios gerais de projeto de um experimento foram explorados da seguinte forma:

- **Aleatorização:** Neste estudo, cada usuário (sujeito) foi atribuído aleatoriamente para cada tratamento (SP ou Leg). Isso implica que cada sujeito avaliou apenas uma versão de cada um dos conteúdos, evitando influências da outra versão. Além disso, todos os objetos (conteúdos multimídia) foram avaliados por todos os usuários e a ordem de apresentação dos conteúdos foi a mesma para todos os usuários.
- **Balanceamento:** O experimento foi balanceado, uma vez que cada tratamento foi avaliado pelo mesmo número de usuários.
- **Bloqueio:** Nenhuma abordagem sistemática de bloqueio foi aplicada.

O teste estatístico aplicado foi o Teste t de *Student* [Jain 1991], um dos testes paramétricos mais utilizados na literatura científica para comparar dois tratamentos.

Instrumentação

De acordo com uma pesquisa realizada pela Associação Brasileira de Rádio e Televisão (ABERT) e pela Fundação Getúlio Vargas (FGV) [FGV/ABERT 2012], a TV Brasileira possui uma programação bastante diversificada envolvendo conteúdos de diversos

gêneros que são distribuídos da seguinte forma: 35,3% da programação são filmes, seriados e novelas; 20,3% são programas jornalísticos; 14,1% são programas infantis; 12,5% são programas de variedade; 5,0% são programas esportivos; 2,6% são programas educacionais; 2,5% são programas de comédia; 2,0% são programas religiosos; 1,4% são *reality shows*, etc.

Com base nessa pesquisa, os conteúdos multimídia selecionados como objetos deste experimento foram quatro conteúdos multimídia que pertencem a diversos gêneros, priorizando os gêneros mais representativos da TV Brasileira. Na Tabela D.1, são apresentadas os conteúdos multimídia que foram utilizados nos testes.

Tabela D.1: Conteúdos multimídia utilizados nos testes

Vídeo	Gênero	Duração	Descrição
Vídeo1	Filmes, seriados e novelas	65 seg.	Esse vídeo é parte de um filme produzido pela TV UFPB, desenvolvido com propósitos acadêmicos.
Vídeo2	Noticiário	26 seg.	Esse vídeo é parte de um programa de notícias veiculado em 14 de Outubro de 2008 na TV Globo, uma rede de TV Brasileira
Vídeo3	Variedades	70 seg.	Esse vídeo é parte de um programa de variedades veiculado em 10 de Novembro de 2011 na TV Record, uma rede de TV Brasileira
Vídeo4	Infantil, filmes	888 seg.	Esse vídeo é um filme curta-metragem de animação 3D produzido pela Blender Foundation ¹

Conforme pode ser observado na Tabela D.1, esses conteúdos representam trechos vídeos de noticiários, de filmes, de programas de variedades (auditório) e de programas infantis, que juntos, representam, em média, cerca de 82,2% dos conteúdos veiculados na TV Brasileira. Isso implica que a amostra utilizada no experimento é representativa. Além disso, os conteúdos são, em geral, curtos para evitar que os testes sejam muito longos e os usuários se entediem durante a avaliação, influenciando na avaliação dos conteúdos apresentados.

Os dados dessa avaliação foram coletados através de questionários. O questionário usado era composto de três partes e é apresentado no Apêndice E. Na primeira parte do questionário, o usuário fornece informações pessoais como sexo, idade, grau de escolaridade e nível de conhecimento em Português e LIBRAS, dentre outros. Na segunda parte, ele responde a 16 (dezesesseis) perguntas sobre os conteúdos apresentados para avaliar o

seu nível de compreensão dos conteúdos. Por fim, na terceira parte, ele avalia subjetivamente os conteúdos apresentados com relação ao nível de compreensão, adequação a gramática de LIBRAS, naturalidade e qualidade da apresentação.

Nas questões utilizadas para verificar se os usuários compreenderam os conteúdos (nível de compreensão), os usuários devem selecionar qual das quatro alternativas (A, B, C ou D) está relacionada ao conteúdo apresentado, onde apenas uma das alternativas é correta. Para todas as questões, a quarta alternativa (D) representava a opção "Não sei responder", que foi incluída para evitar que os usuários escolhessem aleatoriamente uma das alternativas quando não soubessem a resposta correta. Os usuários foram orientados a selecionar esta alternativa quando desconhecessem a resposta correta.

Nas cinco questões da terceira parte, o usuário avaliava os conteúdos com relação ao nível de compreensão, adequação a gramática de LIBRAS (qualidade de tradução), qualidade e naturalidade da apresentação, dentre outros, utilizando uma escala de 1 a 6. Uma escala par (de 1 a 6) foi escolhida porque, de acordo com Morrissey (2008), ela estimula os usuários a avaliarem positivamente ou negativamente os fatores, evitando avaliações neutras. Além disso, essa escala também foi utilizada em trabalhos anteriores relacionados a tradução automática para línguas de sinais, como, por exemplo, em Sansegundo et al. (2011).

Na versão do questionário apresentada para os surdos, todas as perguntas foram traduzidas para glosa por um intérprete de LIBRAS. Essa tradução foi realizada com o objetivo de facilitar a compreensão dos usuários. Adicionalmente, intérpretes de LIBRAS estiveram presentes durante todo o experimento para intermediar a comunicação com os usuários.

D.2.2 Avaliação do Desempenho de Tradução

Seleção do Contexto

Essa parte do experimento foi conduzida através de um conjunto de testes computacionais automáticos desenvolvidos para avaliar o desempenho de tradução da solução proposta considerando um cenário que exige tradução em tempo real (TV). Nesses testes, o LibrasTV, protótipo da solução proposta desenvolvido para TV Digital, foi integrado no ambiente da TV Digital, e utilizou um sinal de uma estação de TV Brasileira como entrada durante um período de 24 (vinte e quatro) horas. Durante esse período, o protótipo processou os conteúdos multimídia transmitidos e o desempenho de tradução foi calculado e armazenado.

Considerando os resultados apresentados pela pesquisa da FGV e ABERT [FGV/ABERT

2012], que indicam que a TV Brasileira possui uma programação bem diversificada, envolvendo programas de diversos gêneros, o teste com um sinal de uma estação de TV Brasileira permitiu avaliar o desempenho da solução para um vocabulário aberto, representativo e em um cenário real.

Formulação das Hipóteses

Informalmente, a hipótese formulada nessa parte do experimento foi que a solução proposta é capaz de traduzir conteúdos multimídia de gêneros variados (domínio geral) em tempo real. Conforme mencionado anteriormente, a solução é considerada de tempo real se ela for capaz de traduzir os conteúdos multimídia com um atraso máximo de quatro segundos (o atraso máximo tolerável para sistemas de transmissão de *closed caption* ao vivo [ABNT 2005]).

Dessa forma, a hipótese dessa parte do experimento foi definida formalmente da seguinte forma:

1. **Hipótese nula H_0 :** A solução proposta não é capaz de traduzir conteúdos com um atraso inferior a quatro segundos, ou seja o tempo médio de tradução é superior a quatro segundos, onde o atraso de tradução é definido como a soma dos tempos de processamento de cada um dos componentes da solução proposta.

$$\mu_{atraso}(SP) > 4segundos \quad (D.5)$$

Hipótese alternativa H_1 : A solução proposta é capaz de traduzir conteúdos com um atraso médio inferior a quatro segundos.

$$\mu_{atraso}(SP) \leq 4segundos \quad (D.6)$$

Seleção das Variáveis

A **variável independente** dessa parte do experimento são os conteúdos multimídia utilizados como entrada para avaliar o desempenho de tradução. Esses conteúdos são extraídos de um sinal de TV Digital real e podem ser de diversos gêneros: filmes, seriados e novelas, programas jornalísticos, programas infantis, programas de variedade, programas esportivos, programas educacionais, programas de comédia, dentre outros.

A **variável dependente** dessa parte do experimento é o tempo (atraso) de tradução dos conteúdos de entrada para LIBRAS.

Instrumentação e Ambiente do Teste

Os conteúdos multimídia selecionados como objetos desse experimento foram extraídos de um sinal real de TV Digital. Para isso o LibrasTV foi integrado no ambiente da TV Digital em um ambiente similar ao apresentado na Seção 5.5.1.

Nesse ambiente, o módulo "Tradutor de LIBRAS" e a aplicação interativa LibrasTV foram executadas em dois mini-computadores (mini-PCs) com processador Intel Dual Core T3200 2GHz e 4GB de memória RAM e sistema operacional Linux Ubuntu 10.0.4, kernel 2.6.32 e foram integrados com um conjunto de equipamentos de TV Digital da seguinte forma:

- Um "Retransmissor IP MPEG-2 TS"² sintonizava o canal de TV Digital e transmitia via *streaming* UDP, o fluxo MPEG-2 TS para o Tradutor de LIBRAS e para o Multiplexador;
- O "Tradutor de LIBRAS" recebia o fluxo MPEG-2 TS do "Retransmissor IP MPEG-2 TS" e gerava um fluxo de LIBRAS codificado com as mensagens LDM e LCM encapsuladas em eventos DSM-CC;
- Um Multiplexador recebia o fluxo MPEG-2 TS e o fluxo de LIBRAS codificado, (re)multiplexava-os em um único fluxo MPEG-2 TS e o transmitia via *streaming* UDP para a máquina do OpenGinga (segundo mini-PC);
- O OpenGinga, com a aplicação LibrasTV embarcada, recebia o fluxo MPEG-2 TS via *streaming* UDP, extraía e executava o vídeo do programa e, em paralelo, executava a aplicação do LibrasTV que gerava as trilhas de LIBRAS a partir do fluxo de LIBRAS codificado.

D.2.3 Avaliação da Ferramenta de Computação Humana

Seleção do Contexto

Essa parte do experimento também foi conduzida na Funad-PB e os sujeitos foram um conjunto de onze surdos brasileiros, três intérpretes de LIBRAS e três *designers-3D* selecionados por conveniência. Os objetos foram um conjunto de cinco sinais em LIBRAS pré-definidos, selecionados de forma que todos os tipos mais comuns de trajetórias de movimentos definidos na LDS sejam contemplados.

Todos os usuários surdos e intérpretes de LIBRAS foram convidados a gerar todos os sinais no WikiLIBRAS, e o número de sinais gerados corretamente por cada usuário foi

²O Retransmissor IP MPEG-2 TS utilizado nesse teste foi desenvolvido pela Mopa Embedded Systems e emprestado gentilmente para a realização desse teste.

medido juntamente com o tempo médio gasto para gerar cada um desses sinais. Paralelamente, os *designers-3D* também foram convidados a gerar os mesmos sinais manualmente utilizando a ferramenta de modelagem e animação Blender e o número de sinais gerados corretamente e o tempo médio para gerar esses sinais também foi computado. Adicionalmente, os surdos e intérpretes de LIBRAS também preencheram um formulário avaliando o grau de satisfação no uso da ferramenta.

Formulação das Hipóteses

A hipótese dessa parte do experimento foi que o tempo médio para gerar sinais corretamente é menor quando o WikiLIBRAS é utilizado por usuários surdos e intérpretes de LIBRAS, do que quando esses sinais são geradas manualmente por *designers-3D*.

Formalmente:

1. **Hipótese nula H_0 :** O tempo médio para gerar um sinal de LIBRAS corretamente no WikiLIBRAS (WL) por usuários surdos e intérpretes de LIBRAS é equivalente ao tempo médio que *designers-3D* levam para gerar sinais manualmente (MAN).

$$\mu_{tempo}(WL) = \mu_{tempo}(MAN) \quad (D.7)$$

Hipótese alternativa H_1 : O tempo médio para gerar um sinal de LIBRAS é menor no WikiLIBRAS, isto é, o WikiLIBRAS é mais produtivo na geração de sinais.

$$\mu_{tempo}(WL) < \mu_{tempo}(MAN) \quad (D.8)$$

Seleção das Variáveis

A **variável independente** desse estudo é o tipo de estratégia utilizada para gerar os sinais. Ela é uma variável nominal que pode assumir dois valores: WL, quando os sinais são gerados no WikiLIBRAS ou MAN, quando os sinais são gerados manualmente por *designers-3D*.

As **variáveis dependentes** são o tempo médio para gerar corretamente sinais em LIBRAS e o grau de satisfação dos usuários no uso da ferramenta de colaboração.

Seleção dos Sujeitos

Os usuários (sujeitos) desse experimento foram onze usuários surdos, três intérpretes de LIBRAS e três *designers-3D* voluntários. Os usuários e intérpretes de LIBRAS

surdos possuíam uma diversificada faixa etária e diferentes graus de escolaridade e eram integrantes da Funad-PB. Os surdos eram alunos de cursos de formação em LIBRAS da Funad-PB e os intérpretes de LIBRAS eram professores/instrutores desses cursos. Os *designers*-3D eram pesquisadores do Laboratório de Aplicações em Vídeo Digital (LAViD) da Universidade Federal da Paraíba (UFPB) com boa experiência em *design*, modelagem e animação 3D.

Projeto do Experimento

A definição, hipótese e as variáveis indicam que o projeto desta parte do experimento também deve ser do tipo **um fator com dois tratamentos** (do inglês, *one factor with two treatments*), onde o fator seria a forma como os sinais são gerados e os tratamentos seriam as possíveis escolhas para esse fator: WikiLIBRAS (WL) e manualmente (MAN).

Os princípios gerais de projeto de um experimento foram explorados da seguinte forma:

- **Aleatorização:** Não abordagem sistemática de aleatorização foi aplicada. Todos os objetos (sinais) foram atribuídos a todos os usuários.
- **Balanceamento:** O experimento era balanceado, uma vez que os mesmos objetos foram atribuídos para todos os usuários dos dois tratamentos.
- **Bloqueio:** Nenhuma abordagem sistemática de bloqueio foi aplicada.

O teste estatístico aplicado também foi o Teste t de *Student* [Jain 1991].

Instrumentação

Os objetos do experimento foram os sinais apresentados na Tabela D.2. Conforme pode ser observado nessa tabela, os sinais selecionados abrangem os tipos de movimentos mais comuns definidos na LDS (retilíneo, circular, semi-circular e pontual) [Gibet et al. 2001]. Eles foram selecionados de acordo com o tipo de movimento, porque a interação no WikiLIBRAS é direcionada pelo tipo de movimento (ver Seção 5.4) e, dessa forma, seria possível abranger os diversos tipos de interação da ferramenta. Além disso, o número de sinais utilizados no experimento foi pequeno, para evitar que os testes fossem muito longos e os usuários se entediasssem durante a avaliação, influenciando na avaliação da ferramenta e na qualidade dos resultados.

Os dados dessa avaliação foram coletados através de questionários. O questionário usado nessa avaliação também foi composto de três partes e é apresentado no Apêndice E. Na primeira parte do questionário, são coletadas informações pessoais dos usuários,

Tabela D.2: Sinais utilizados nos testes

Sinal	Tipo de Movimento
PRESIDENTE	Retilíneo
LÁBIO	Circular
PROFESSOR	Semi-Circular
CALAR	Pontual
TIO	Pontual

como, por exemplo, sexo, idade, grau de escolaridade, grau de conhecimento em LIBRAS, dentre outros. Na segunda parte, os usuários avaliam se conseguiram gerar os sinais e indicam quais as principais dificuldades enfrentadas nos sinais que não foram gerados corretamente. Na terceira parte, são avaliados alguns aspectos relacionados ao grau de satisfação dos usuários no uso da ferramenta.

Na versão do questionário apresentada para os surdos, todas as perguntas foram traduzidas para glosa com o objetivo de facilitar sua compreensão. Adicionalmente, uma intérprete de LIBRAS esteve presente durante todo o experimento para intermediar a comunicação com os surdos.

Apêndice E

Questionários

E.1 Questionário 1: Avaliação dos Conteúdos Acessíveis

Parte 1: Informações dos Usuários

Sexo: Masculino Feminino

Idade: _____

Você sabe ler? Sim Não

Qual é o seu grau de escolaridade?

- Nunca estudei
- Ensino Fundamental Incompleto
- Ensino Fundamental Completo
- Ensino Médio Incompleto
- Ensino Médio Completo
- Ensino Superior Incompleto
- Ensino Superior Completo
- Outro. Qual? _____

Qual o seu nível de conhecimento em Português (6 - avançado, 1 - não sei nada)?

- 6 5 4 3 2 1

Qual o seu nível de conhecimento em LIBRAS (6 - avançado, 1 - não sei nada)?

- 6 5 4 3 2 1

Parte 2: Compreensão dos Conteúdos**Vídeo1:**

- 1) O vídeo mostra um casal discutindo. Qual é o motivo da discussão?
 - A) Ciúmes
 - B) Problemas financeiros
 - C) Problemas com os filhos
 - D) Não sei responder

- 2) Por que a discussão do casal começou?
 - A) Porque só tinha café
 - B) Porque a esposa tinha chegado de madrugada no dia anterior
 - C) Porque os filhos não acordaram para ir a escola
 - D) Não sei responder

- 3) Onde o marido mandou a esposa resolver o problema?
 - A) No mercado
 - B) No médico
 - C) No advogado
 - D) Não sei responder

- 4) Qual era o motivo que a esposa alegava que não podia ir resolver o problema?
 - A) Porque o médico não podia atender
 - B) Porque o dono do mercado não vendia fiado
 - C) Porque o advogado era caro
 - D) Não sei responder

- 5) No final do vídeo, o marido, chateado, diz o que para a esposa?
 - A) "Nossos filhos não podem faltar a escola"
 - B) "Você tem que dar um jeito de ir ao médico"
 - C) "O que não pode é sair para trabalhar sem comer"
 - D) Não sei responder

Vídeo2:

1) Qual foi a catástrofe apresentada na reportagem?

- A) Terremoto (tremor de terra)
- B) Tempestade (muita chuva)
- C) Tsunami (onda gigante)
- D) Não sei responder

2) Quantos corpos foram encontrados soterrados?

- A) 1
- B) 2
- C) 5
- D) Não sei responder

3) O que aconteceu com a casa dos moradores da região?

- A) A casa pegou fogo
- B) Outros moradores se mudaram para lá
- C) A chuva derrubou a casa
- D) Não sei responder

Vídeo3:

1) As apresentadoras discutem sobre o que no início do vídeo?

- A) Economia
- B) Política
- C) Moda
- D) Não sei responder

2) Qual foi a recomendação da apresentadora Gianne Albertoni?

- A) Não comam muita massa
- B) Não usem muito brilho
- C) Usem muito brilho
- D) Não sei responder

3) Qual é a receita apresentada por Edu Guedes?

- A) Doce de Pêssego

- B) Doce de Mamão
- C) Mousse de Maracujá
- D) Não sei responder

4) A Melissa que apresentou a receita com o Edu Guedes possui uma outra ocupação além de cozinhar. Que ocupação é essa?

- A) Melissa também é professora
- B) Melissa também trabalha com animais
- C) Melissa também vende roupas
- D) Não sei responder

Vídeo4:

1) Qual é o animal de estimação da garota (personagem da história)?

- A) Galinha
- B) Gato
- C) Dragão
- D) Não sei responder

2) O que o guardião disse a garota?

- A) Que ela estava próxima de encontrar seu animal perdido
- B) Que ela estava longe de encontrar o seu animal perdido
- C) Que ela era filha dele
- D) Não sei responder

3) Qual é a história da lança encontrada pela garota?

- A) Ela tinha um passado triste, havia assassinado muitos inocentes.
- B) Ela tinha um passado glorioso, havia assassinado muitas pessoas más
- C) Ela pertenceu a Hércules, um grande guerreiro
- D) Não sei responder

4) O que significa Sintel?

- A) Terra perigosa
- B) Terra fria
- C) Terra dos dragões
- D) Não sei responder

Parte 3: Avaliação dos Conteúdos

1) É fácil de entender? (6 - Claro, 1 - Confuso)

6 5 4 3 2 1

2) Seguem a gramática de LIBRAS? (6 - Perfeita, 1 - Inadequada)

6 5 4 3 2 1

3) A sinalização é natural? (6 - Parece uma pessoa, 1 - Parece um robô)

6 5 4 3 2 1

4) Os movimentos do avatar-3D estão corretos? (6 - Perfeitos, 1 - Inadequados)

6 5 4 3 2 1

5) As expressões faciais estão corretas? (6 - Perfeitas, 1 - Inadequadas)

6 5 4 3 2 1

E.2 Questionário 2: Avaliação do WikiLIBRAS

Parte 1: Informações dos Usuários

Sexo: Masculino Feminino

Idade: _____

Você é? Surdo Intérprete de LIBRAS

Você sabe ler? Sim Não

Qual é o seu grau de escolaridade?

Nunca estudei

Ensino Fundamental Incompleto

Ensino Fundamental Completo

Ensino Médio Incompleto

Ensino Médio Completo

Ensino Superior Incompleto

Ensino Superior Completo

Outro. Qual? _____

Você utiliza internet?

Nunca usei Quase nunca Pouco Frequentemente Todos os dias

O que você mais acessa quando utiliza internet? (Pode marcar mais de uma alternativa)

Sites de Busca (Ex: Google, Yahoo, Bing)

Redes Sociais (Ex: Orkut, Facebook, Twitter)

E-mail

Bate-papo (Ex: MSN, Google Talk)

Notícias (Ex: Uol, G1)

Jogos (Ex: Warcraft, Starcraft, Diablo)

Vídeos (Ex: Youtube, GoogleVideo)

Outro. Qual? _____

Qual o seu nível de conhecimento em Português (6 - avançado, 1 - não sei nada)?

6 5 4 3 2 1

Qual o seu nível de conhecimento em LIBRAS (6 - avançado, 1 - não sei nada)?

6 5 4 3 2 1

Parte 2: Geração dos Sinais

1) Agora, tente, por gentileza, gerar os sinais abaixo no WikiLIBRAS e indique se você conseguiu ou não gerar cada um dos sinais?

Sinal PRESIDENTE. Conseguiu gerar corretamente? Sim Não

Sinal PROFESSOR. Conseguiu gerar corretamente? Sim Não

Sinal CALAR. Conseguiu gerar corretamente? Sim Não

Sinal TIO. Conseguiu gerar corretamente? Sim Não

Sinal LÁBIO. Conseguiu gerar corretamente? Sim Não

2) Para os sinais que você não conseguiu gerar, indique qual(is) foi(ram) a(s) maior(es) dificuldade(s):

Sinal PRESIDENTE

- Não entendi como utilizar o site
- Não entendi os termos utilizados no site
- O site não funcionou corretamente
- Não sei representar esse sinal em LIBRAS

Sinal PROFESSOR

- Não entendi como utilizar o site
- Não entendi os termos utilizados no site
- O site não funcionou corretamente
- Não sei representar esse sinal em LIBRAS

Sinal CALAR

- Não entendi como utilizar o site
- Não entendi os termos utilizados no site
- O site não funcionou corretamente
- Não sei representar esse sinal em LIBRAS

Sinal TIO

- Não entendi como utilizar o site
- Não entendi os termos utilizados no site
- O site não funcionou corretamente
- Não sei representar esse sinal em LIBRAS

Sinal LÁBIO

- Não entendi como utilizar o site
- Não entendi os termos utilizados no site
- O site não funcionou corretamente
- Não sei representar esse sinal em LIBRAS

Parte 3: Avaliação da Ferramenta

1) O sistema é fácil de usar? (6 - Muito fácil, 1 - Muito difícil)

6 5 4 3 2 1

2) A sinalização do avatar-3D parece natural? (6 - Parece uma pessoa, 1 - Parece um

robô)

6 5 4 3 2 1

3) É fácil configurar a mão dos sinais? ? (6 - Muito fácil, 1 - Muito difícil)

6 5 4 3 2 1

4) É fácil configurar o ponto de articulação dos sinais? (6 - Muito fácil, 1 - Muito difícil)

6 5 4 3 2 1

5) É fácil configurar a orientação da palma da mão dos sinais? (6 - Muito fácil, 1 - Muito difícil)

6 5 4 3 2 1

6) É fácil configurar a expressão facial dos sinais? (6 - Muito fácil, 1 - Muito difícil)

6 5 4 3 2 1

