Centro Federal de Educação Tecnológica de Minas Gerais
Programa de Pós-graduação em Modelagem Matemática e Computacional

Abordagem Computacional para Criação de Neologismos Terminológicos em Línguas de Sinais

Celso Luiz de Souza

Orientador: Prof. Dr. Flávio Luis Cardeal Pádua
Centro Federal de Educação Tecnológica de Minas Gerais

Coorientadora: Profa. Dra. Vera Lúcia de Souza e Lima
Centro Federal de Educação Tecnológica de Minas Gerais

Belo Horizonte, MG
Março de 2018

Celso Luiz de Souza

Abordagem Computacional para Criação de Neologismos Terminológicos em Línguas de Sinais

Tese apresentada ao Programa de Pós-graduação em Modelagem Matemática e Computacional do Centro Federal de Educação Tecnológica de Minas Gerais, como requisito parcial para a obtenção do título de Doutor em Modelagem Matemática e Computacional.

Área de concentração: Modelagem Matemática e Computacional

Linha de pesquisa: Métodos Matemáticos Aplicados

Orientador: Prof. Dr. Flávio Luis Cardeal Pádua
Centro Federal de Educação Tecnológica de Minas Gerais

Coorientadora: Profa. Dra. Vera Lúcia de Souza e Lima
Centro Federal de Educação Tecnológica de Minas Gerais

Centro Federal de Educação Tecnológica de Minas Gerais
Programa de Pós-graduação em Modelagem Matemática e Computacional
Belo Horizonte, MG
Março de 2018

Souza, Celso Luiz de
S729a Abordagem computacional para criação de neologismos terminológicos em línguas de sinais / Celso Luiz de Souza. – 2018.
84 f.

Tese de doutorado apresentada ao Programa de Pós-Graduação em Modelagem Matemática e Computacional.
Orientador: Flávio Luis Cardeal Pádua.
Coorientadora: Vera Lúcia de Souza e Lima.
Tese (doutorado) – Centro Federal de Educação Tecnológica de Minas Gerais.

1. Língua de sinais – Processamento de dados – Teses. 2. Iconicidade (linguística) – Teses. 3. Língua brasileira de sinais – Teses. I. Pádua, Flávio Luis Cardeal. II. Lima, Vera Lúcia de Souza e. III. Centro Federal de Educação Tecnológica de Minas Gerais. IV. Título.

CDD 004.019

Elaboração da ficha catalográfica pela Biblioteca-Campus II / CEFET-MG

Dedico primeiramente este trabalho às pessoas que necessitam de apoio da comunidade científica para o seu desenvolvimento.

Ao meu primo Luiz Carlos Nogueira Júnior, no qual sempre me inspirei tanto na vida acadêmica quanto na profissional, a este agradeço enormemente o seu apoio.

A minha família, em especial minha esposa, Ana Cristina, pela dedicação e pela confiança em meu trabalho.

Agradecimentos

Ao meu orientador, Flávio Luis Cardeal Pádua: serei eternamente grato pela oportunidade de realizar este trabalho. À minha coorientadora Vera Lúcia de Souza e Lima: pelo seu apoio e sua disponibilidade para o desenvolvimento deste trabalho.

Aos amigos de trabalho, Carlos Carneiro e Anísio Lacerda, que ajudaram no desenvolvimento da minha tese.

Aos meus pais, Celso Souza e Maria das Mercês Teixeira Souza pelas oportunidades que me ofereceram.

Não teria como deixar de agradecer ao meu primo e "irmão", Luiz Carlos Nogueira Junior, que sempre foi um grande parceiro e incentivador em minha carreira acadêmica, me apoiou em vários momentos e sempre disse: "tudo é persistência"!

Aos meus colegas de trabalho e amigos que ajudaram na compreensão da língua de sinais, Kate Lyra, Gilberto Goulart, Felipe D. Castro, Raquel Barreto, Madson Barreto, Ana Rachel e todos os participantes da pesquisa que foi desenvolvida.

Às minhas filhas Carla e Bruna que me inspiraram a continuar meus estudos, e minha amada Ana Cristina que sempre está a meu lado apoiando e auxiliando com suas revisões valiosas.

Agradeço também as instituições CEFET-MG pela oportunidade e IF Sudeste MG pelo afastamento imprescindível em minha qualificação.

"Talvez não tenha conseguido fazer o melhor, mas lutei para que o melhor fosse feito. Não sou o que deveria ser, mas Graças a Deus, não sou o que era antes." (Marthin Luther King)

Resumo

O número de surdos que estão ativamente no meio acadêmico está aumentando, os quais deparam-se com diversos tipos de barreiras que dificultam o seu desenvolvimento. Uma dessas barreiras está associada à escassez de um léxico específico para termos técnicos na língua de sinais em que têm proficiência. Visando-se solucionar esse problema em particular, a presente tese apresenta uma abordagem inovadora para criação de neologismos terminológicos (sinais técnicos) para as áreas de Ciência, Tecnologia, Engenharia e Matemática (CTEM) em língua de sinais. A abordagem proposta baseia-se na utilização de métodos computacionais para auxiliar o processo de criação de neologismos terminológicos que constituirão novos dicionários em CTEM. Especificamente, propõe-se: (1) a extração de descritores de forma a partir da imagem contendo a ilustração do significado do conceito técnico sem sinal de interesse, utilizando-se visão computacional, e (2) a determinação do radical do termo associado ao conceito na língua oral de referência do surdo, utilizando-se processamento de linguagem natural. Ao comparar essas informações do conceito técnico sem sinal com as correspondentes de sinais de propósito geral em um repositório, a presente abordagem cria um novo sinal semanticamente aceitável a partir de sinais pré-existentes com características semelhantes. No estado da arte atual, não foram identificados trabalhos cujas soluções para criação de sinais sejam apoiadas por métodos computacionais. Dito isso, a abordagem proposta nesta tese é potencialmente a primeira na literatura a apoiar a produção de novos sinais com o auxílio da computação, tornando este processo mais parametrizado, ágil e escalável. Para avaliar a qualidade e a utilidade dos sinais técnicos criados por meio desta abordagem, considerou-se neste trabalho um estudo de caso no âmbito da Língua Brasileira de Sinais (Libras). Em particular, foram avaliados sinais em Libras criados para conceitos técnicos da área de Desenho Arquitetônico. Constituiu-se, para este fim, um comitê avaliador composto por falantes de Libras e conhecedores de escrita de sinais, os quais realizaram suas avaliações tomando como base aspectos fonéticos, fonológicos e a própria estrutura morfológica da língua de sinais em questão. Os resultados obtidos sugerem que a abordagem proposta é promissora, sendo que cerca de 60% dos novos sinais criados foram classificados com nível de satisfação similar àquele atribuído a sinais de uma base de dados de referência (criados manualmente), em uma escala de classificação de cinco níveis.

Palavras-chave: Língua de sinais. iconicidade. Libras.

Abstract

The number of people with hearing impairment who are actively in academia is increasing and they face several types of barriers which make harder their development. One of those barriers is associated with the scarcity of a specific lexicon for technical terms in the sign language which they are proficient. Aiming to settle this particular issue, this thesis presents an innovating approach to create terminological neologisms (technical signs) for Science, Technology, Engineering and Mathematics (STEM) areas in sign language. The proposed approach is based on using of computational methods to support the creation of terminological neologisms which will compose new dictionaries of STEM. It is specifically proposed that: (1) shape descriptors will be extracted from an image containing the illustration of the technical meaning without any sign of interest, using computer vision, and (2) determination of the stemming of the term associated with the reference concept in oral language for people with hearing impairment, using natural language processing. When the information about technical concept without sign is compared to its correspondent with signs of general purpose in a repository, this approach creates a new sign which is semantically acceptable originated from preexistent signs with similar characteristics. Nowadays, there are not any identified researches in this field in which computational methods were used as a solution to create signs. So, the approach related in this thesis is potentially the first one in literature about production of new signs through computing support, making the process more parameterized, agile and scalable. To evaluate the quality and the utility of the technical signs created through this approach, a case study concerning the Brazilian Sign Language (Libras) was used. Particularly, signs from Libras that were created for technical concepts of architectural design were evaluated. A committee, composed by Libras speakers and connoisseurs of the written sign language, was responsible to evaluate it basing on aspects of phonetic, phonology and the morphology of the considered sign language. The obtained results pointed that it is a promising purpose, because about 60% of the new created signs were classified with a similar level of satisfaction of the ones (manually created) attributed to signs from a database reference, whose classification consists in five levels.

Keywords: sign language. iconic. Libras.

Lista de Figuras

Figura 1 – Dicionário de Libras que utiliza vídeo para apresentar o sinal. 4
Figura 2 – Dicionário de Libras, que utiliza animação para apresentar o sinal. 4
Figura 3 – Dicionários em Língua de Sinais, para vários países. 5
Figura 4 – A apresentação de conceitos na sala de aula tradicionais. 6
Figura 5 – Há grande possibilidade de criação de sinais diferentes para o mesmo conteúdo, em diferentes instituições de ensino, visto que frequentemente os trabalhos de criação de tais sinais são feitos de forma independente e sem considerar regras ou critérios específicos. 6
Figura 6 – Uma visão da abordagem proposta para geração de novos sinais técnicos para língua de sinais. 9
Figura 7 – Ilustração do termo "CASA": 7a em Libras e respectivamente em português; 7b datilologia do termo e as letras alfabéticas. 12
Figura 8 – Sinais não manuais dos termos "LADRÃO": 8a em Libras e respectivamente em português; 8b Libras e respectivamente em português. 14
Figura 9 – Locação do sinal - espaço do sinal. 16
Figura 10 – Processo de derivação dos nomes dos verbos. 17
Figura 11 – Processo de composição. 18
Figura 12 – Diagrama que ilustra a proposta de Saussure. 19
Figura 13 – Configurações de mãos propostas por Stokoe. 21
Figura 14 – Ilustração da escrita em SW da palavra **casa**. 22
Figura 15 – Exemplo da linguagem SWML. 23
Figura 16 – SingBox - Layout de posicionamento dos símbolos SW. 25
Figura 17 – Ilustração do descritor e da correspondência do *Shape Context* entre imagens, as Figuras: 17a, 17e são as imagens originais; 17b contorno das bordas; 17f são os vetores de um ponto; 17c é o histograma dos vetores da Figura 17f; 17g é as correspondências; 17d e 17h são os resultados dos processamentos. 31
Figura 18 – Fluxograma do algoritmo de Orengo e Huyck (2001). 33
Figura 19 – Visão geral sobre as áreas a serem exploradas. 34
Figura 20 – Gráfico da análise do processo de formação de sinais. 47
Figura 21 – Visão geral do abordagem proposta. 50
Figura 22 – Para criação de um novo sinal em Libras para o conceito "fachada", considera-se como entrada o termo "fachada" e a sua ilustração. 51
Figura 23 – Exemplo das operações da função de Levenshtein. 54
Figura 24 – Dois exemplos da fusão das classificações, considerando as descrições das imagens e termos. Há quatro sinais, nomeados $\{A, B, C, D\}$, a partir dos quais será selecionado o modelo do sinal. 56

Figura 25 – Desempenho da abordagem proposta da métrica DMOS para cada conceito de interesse. 65

Lista de Tabelas

Tabela 1 – Definição das marcações em SW. 24
Tabela 2 – Definição da expressão regular do FSW. 26
Tabela 3 – Número de processos linguísticos envolvidos na formação dos sinais propostos no trabalho de Lima (2014). 46
Tabela 4 – Resumos das principais obras que desenvolveram glossários gerais e específicos. 48
Tabela 5 – Tabela de termos e seus métodos de formação. 61
Tabela 6 – Análise dos processo de formação de Sinais. 63
Tabela 7 – Resumo dos resultados das métricas. 65
Tabela 8 – Resultado dos testes: Wilcoxon (TW) e Sinais (TS). N = número de amostras, Z valor da estatística do teste e p é probabilidade associada. Os valores em negrito mostram diferenças estatisticamente significativas. . . . 66
Tabela 9 – Os dez sinais propostos pelas abordagens CMS e CSAC para cada conceito técnico do desenho arquitetônico. 68

Lista de Quadros

Quadro 1 – Diferenças entre produções na língua portuguesa e em Libras. 13

Lista de Abreviaturas e Siglas

ACR-HR	*Absolute Category Rating with Hidden Reference*
ASL	*American Sign Language*
BSL	*British Sign Language*
CEP	Comitê de Ética em Pesquisa
CMS	Criação Manual de Sinais
CSAC	Criação de Sinais Assistida por Computador
CTEM	Ciência, Tecnologia, Engenharias e Matemática
DGS	*Deutsche Gebärdensprache*
DMOS	*Diference Mean Opinion Score*
ID	Identificador
Libras	Língua Brasileira de Sinais
MOS	*Mean Opinion Score*
SLN	*Sign Language of the Netherlands*
SMD	*Signing Math Dictionary*
SSC	*Scottish Sensory Centre*
STABOL	*Standardization of Basic Lexicon*
STEM	*Science, Technology, Engineering, and Mathematics*
TS	Teste de Sinais
TSW	Teste dos Sinais de Wilcoxon
UDL	*Universal Design for Learning*

Sumário

1 – Introdução 1
1.1 Definição do Problema de Pesquisa 5
1.2 Motivação 7
1.3 Objetivos: Geral e Específicos 8
1.4 Contribuições 10
1.5 Organização do trabalho 10

2 – Fundamentação Teórica 11
2.1 Línguas de Sinais 11
2.1.1 As relações entre Línguas de Sinais e Línguas Orais 12
2.1.2 Sinais Manuais e Não Manuais 14
2.1.3 Parâmetros Fonéticos e Fonológicos das Línguas de Sinais 15
2.1.3.1 Restrições na Formação de Sinais 16
2.1.3.2 Criação de Sinais 17
2.1.3.2.1 Morfologia 17
2.1.3.2.2 Noções sobre Arbitrariedade e Iconicidade 19
2.1.4 Escrita de Sinais 20
2.1.4.1 *Sign Writing* 22
2.1.5 Língua Brasileira de Sinais (Libras) 26
2.1.5.1 Dicionários em Libras 27
2.1.5.1.1 Dicionários Terminológicos em Libras 28
2.2 Educação de Surdos e Novas Tecnologias 28
2.3 Visão Computacional 30
2.3.1 Descritor de Forma 30
2.3.1.1 *Shape Context* 30
2.4 Processamento de Linguagem Natural 32
2.4.1 Extração de Radicais 32

3 – Trabalhos Relacionados 34
3.1 Educação inclusiva 35
3.2 Dicionários para língua de sinais 36
3.2.1 Dicionários de Propósito Geral 37
3.2.2 Dicionários específicos 38
3.3 Formação das palavras 40
3.3.1 Morfologia tradicional 40
3.3.2 Iconicidade 42

3.3.3 Arbitrariedade . . . 44
3.3.4 Demais Processos Linguísticos na Formação de Sinais . . . 46
3.4 Discussão . . . 47

4 – Abordagem Proposta . . . 50
4.1 Classificação por Ilustração . . . 51
4.2 Classificação por Radical . . . 52
4.3 Fusão das Classificações . . . 54
4.4 Busca pelo símbolo modelo . . . 56
4.5 Composição do novo sinal . . . 57

5 – Resultados experimentais . . . 58
5.1 Ambiente de experimentos . . . 58
5.1.1 Repositório de Dados . . . 60
5.1.2 Questionário . . . 61
5.1.3 Métricas . . . 63
5.1.4 Reprodutibilidade . . . 64
5.2 Discussões . . . 64
5.2.1 Satisfação da proposta . . . 64
5.2.2 Associação entre sinal e significado . . . 67
5.2.3 Analise qualitativa . . . 67

6 – Conclusão . . . 69
6.1 Resultados alcançados . . . 70
6.2 Trabalhos futuros . . . 70

Referências . . . 72

Apêndices 83

APÊNDICE A – Parecer do CEP . . . 84

Capítulo 1

Introdução

A presente tese propõe uma abordagem computacional, baseada em métodos de visão computacional e processamento de linguagem natural para apoiar o desenvolvimento de neologismos terminológicos relativos a conceitos técnicos inexistentes no léxico ou no conjunto de palavras de uma dada língua de sinais. Entende-se que o sinal em uma dada língua de sinais é o mesmo que uma palavra na língua oral, por exemplo. Esse conjunto de palavras ou de terminologias é usualmente abordado em disciplinas de cursos das áreas de Ciência, Tecnologia, Engenharia e Matemática (CTEM). De fato, a escassez de sinais ou de termos em CTEM tem preocupado estudantes surdos, intérpretes da Língua Brasileira de Sinais (Libras) e professores de várias disciplinas acerca da assimilação dos conceitos veiculados em salas de aula uma vez que os vocabulários técnicos em CTEM não são encontrados nos vernáculos das principais línguas de sinais do mundo.

Sabe-se que, neste contexto, intérpretes, professores e tutores esforçam-se para desenvolver ou criar novas terminologias em tais áreas, frequentemente por meio de projetos e de organizações locais, os quais, em alguns casos são compostos adicionalmente por pessoas da comunidade surda diretamente interessadas nestas novas terminologias. Na maioria das vezes, trata-se de um trabalho complexo, de longa duração, frequentemente sem uma metodologia clara e, portanto, de difícil escalabilidade. Além disso, por meio deste tipo de abordagem, verifica-se que sinais completamente distintos entre si podem ser atribuídos a um mesmo conceito técnico, quando intérpretes ou instituições de ensino diferentes realizam esse trabalho de forma independente. Embora, principalmente no ambiente escolar, terminologias específicas de diversas áreas em CTEM estejam sendo pensadas e criadas de fato, faltam registros sistematizados desses sinais que abordem conceitos importantes relacionados aos princípios da terminografia (LIMA, 2014).

Portanto, nota-se que o problema da escassez de um léxico específico para termos técnicos é algo crítico e de difícil solução, sobretudo quando se considera que os vocabulários usados em CTEM trazem consigo desafios diferentes daqueles observados em situações de leitura comuns, não apenas para pessoas com perda auditiva, mas para todos os alunos de maneira geral.

Ademais, deve-se ressaltar que a maioria dos indivíduos surdos ou com algum grau severo de deficiência auditiva possuem uma grande dificuldade em compreender a língua oral do país em que vivem. Ou seja, tais indivíduos são fortemente dependentes da interpretação da língua de sinais que utilizam uma vez que a compressão quanto ao que está escrito é quase inexistente.

Adicionalmente, pessoas com deficiência auditiva ou surdas sofrem exclusão e preconceito e são frequentemente consideradas "não normais" pela sociedade ouvinte (QUADROS; PERLIN, 2007; RODRIGUES; MUNHOZ; HATTGE, 2015). O acesso do surdo ao ambiente escolar é muitas vezes conturbado. Desde o ambiente residencial até ao escolar são inúmeras barreiras a serem transpostas, e, de fato, estas barreiras muitas vezes fazem com que estes indivíduos se afastem do estudos. A sociedade muitas vezes buscou a pseudo "cura" da surdez, impondo a esses indivíduos práticas e tratamentos médicos traumatizantes, como, as cirurgias de implantes cocleares, que são consideradas uma violência pelas comunidades surdas (STUMPF, 2005). Esse preconceito faz com que os surdos fiquem excluídos das instituições de ensino por não terem acesso ao aprendizado de maneira semelhante aos não surdos.

Compreende-se que uma das formas de aprendizado dos alunos é realizada através da leitura e da escrita (FERREIRA; DIAS, 2002). O indivíduo, em seu aprendizado, depara-se com novos conceitos, que são usados recorrentemente nos universos desconhecidos que ele passa a explorar. A assimilação desses conceitos inéditos faz-se por meio das associações que o indivíduo acredita serem pertinentes aos universos conceituais em que é introduzido. Uma das formas de associação mais comum dá-se através do uso da escrita. A escrita permite, além da disseminação do conhecimento, a preservação do mesmo. Conforme sentencia o velho provérbio latino, *Verba volant, scripta manent*[1], ou seja, ideias e conceitos novos transmitidos verbalmente entre um grupo de pessoas podem perder-se no tempo. Por outro lado, a escrita é capaz de perpetuar a história, os conceitos, as ideias e os conhecimentos.

Levando em consideração esses apontamentos, mostra-se necessário analisar os processos de assimilação e de aprendizagem, quando vislumbrados a partir da experiência das pessoas surdas. Os indivíduos surdos começam o aprendizado enfrentando uma enorme barreira, devido ao modelo educacional imposto pela comunidade não surda, o qual tem como base o uso da língua oral. A conexão entre linguagem, cognição e aprendizagem para indivíduos surdos tem sido amplamente estudada nos últimos anos (MARSCHARK; HAUSER, 2008) com o objetivo de auxiliá-los no aprendizado. Um dos principais assuntos discutidos pelos pesquisadores é o aspecto cognitivo. Atenção especial tem sido dedicada ao fato de que os alunos surdos são menos propensos a inferir e a relacionar conceitos ao longo de várias dimensões, habilidades essas necessárias para a compreensão do estudo da ciência e, consequentemente, de outras áreas, como tecnologia e engenharia (MARSCHARK; HAUSER, 2008; LANG et al., 2007). Esses estudos indicam, em seus resultados, a necessidade de uma abordagem sistemática para

[1] "As palavras voam, a escrita permanece."

examinar tanto a seleção dos sinais, quanto os métodos de criação de sinais e seu impacto sobre a vida acadêmica dos alunos surdos. Diferentes autores argumentam que a falta de um léxico da língua de sinais seja para área científica ou técnica pode prejudicar o envolvimento dos alunos surdos no processo de aprendizagem (REDDEN; DAVIS; BROWN, 1978; LANG; EGELSTON-DODD; SACHS, 1983; LANG, 2002; LANG et al., 2007; LIMA, 2014).

Alguns pesquisadores começaram a tratar dos assuntos relacionados à linguística na Libras recentemente. Pizzio (2011) relata que tudo se iniciou na década de 1980, tendo como foco aspectos relacionados à fonologia e à sintaxe, ressaltando ainda que são raros os trabalhos voltados à morfologia e à fonética e que quando existem são superficiais e se baseiam na língua de sinais americana, o que pode ocasionar relações errôneas ao tratar de outra cultura.

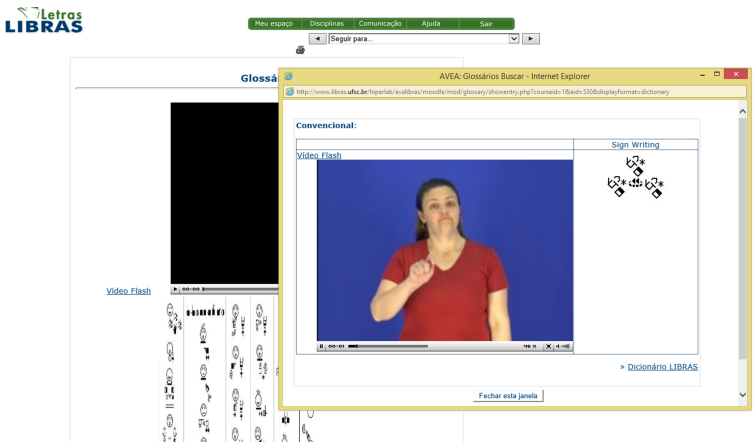
Além disso, a educação dos surdos atualmente passa pela necessidade de um intérprete, que deve ter conhecimento sobre o conteúdo a ser repassado. Segundo Dias, Pedroso e Rocha (2006), o êxito no processo de aprendizagem dos alunos surdos está associado ao uso da língua de sinais como língua materna. Stadler, Filietaz e Hussein (2012) observa que a prática bilíngue no ensino de crianças com deficiência auditiva é recomendada por contribuir no processo de aprendizado, porém não é eficaz.

Considerando estes aspectos, seja cultural, da escrita ou do conhecimento em relação ao conceito, alguns projetos propuseram o desenvolvimento de glossários ou de dicionários em Libras. Em geral, na tipologia de dicionários, são considerados como de Língua Geral, tal como aquele proposto por Capovilla e Raphael (2001a), Capovilla e Raphael (2001b), o qual é impresso.

Tais dicionários começaram a preencher uma lacuna na ausência de dicionários específicos para os surdos. Alguns projetos desenvolveram dicionários gerais em que o acesso é realizado pela Internet. Esses sistemas são voltados para os falantes da língua oral, ao assumir as buscas por palavras em Português por exemplo, escritas de forma arábica, e a representação do sinal em vídeo, como, por exemplo, o glossário Letras Libras desenvolvido por Stumpf, Oliveira e Miranda (2006), o qual é ilustrado na Figura 1. Outro exemplo interessante é o glossário denominado Dicionário Libras desenvolvido por Marques (2002), que é ilustrado na Figura 2. Finalmente, deve-se mencionar o repositório SignPuddle (SLEVINSKI, 2012), que apresenta uma vasta quantidade de dicionários para vários idiomas, inclusive a Libras, conforme ilustrado na Figura 3. É observado que todos os trabalhos apresentam somente o sinal, não informando nenhum aspecto sobre a sua morfologia, e, quanto aos critérios de inserção do termo na base de dados, estes ocorrem de maneira totalmente livre em alguns casos, não havendo nenhum processo de validação dos termos inseridos.

Pode-se observar adicionalmente que estes sistemas não abordam questões relacionadas à linguística, como a descrição do processo de formação ou a aplicação do termo. Em geral, o

Figura 1 – Dicionário de Libras que utiliza vídeo para apresentar o sinal.



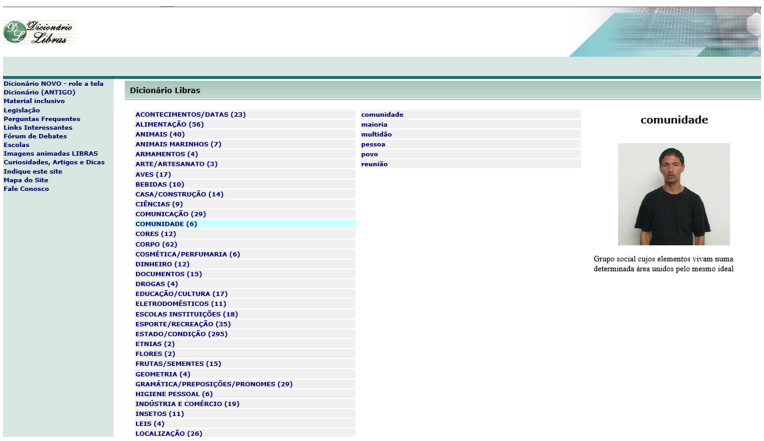
Fonte: Letras Libras - UFSC.

foco principal é meramente tratar os aspectos computacionais do sistema e a apresentação dos sinais, em uma espécie de tradutor de palavras.

Não se pode deixar de relatar que muitos acreditam que a língua de sinais é a comunicação realizada por meio de mímicas, sendo esse um mecanismo rudimentar e sem regras de comunicação, ou em alguns casos acredita-se que há um alfabeto que possa se relacionar diretamente com a palavra. Hoje, esse conceito errôneo já não possui tanta evidência quanto no passado, sendo que os surdos têm hoje a construção de uma identidade a partir de uma língua e de uma cultura próprias, definido por Strobel (2008, p. 27) como a forma que o sujeito surdo compreende e molda o mundo para torna-lo acessível as suas percepções visuais.

Os aspectos característicos da cultura surda devem ser respeitados em todas as dimensões, sejam sociais, políticas e culturais, incluído o direito de uso da Língua de Sinais, no caso do Brasil, a Libras (NASCIMENTO, 2009; STUMPF, 2005). Em suma, as comunidades surdas

Figura 2 – Dicionário de Libras, que utiliza animação para apresentar o sinal.



Fonte: Dicionário Libras.

Figura 3 – Dicionários em Língua de Sinais, para vários países.

Fonte: SignPuddle

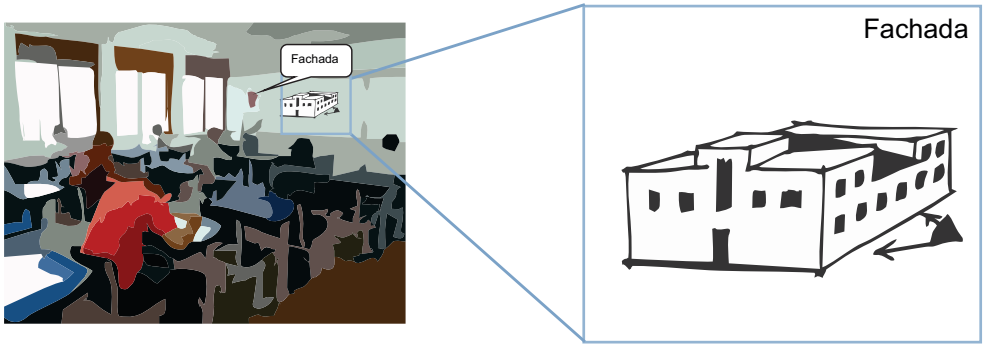
encontram-se em um momento ímpar de lutas por direitos que lhes foram negados durante décadas (GUIMARÃES, 2013).

Diante deste cenário, a presente tese aborda o problema da escassez de léxicos específicos para termos técnicos em línguas de sinais, problema este de natureza interdisciplinar, por meio da proposição de uma solução fundamentada em conceitos e métodos de duas grandes áreas do conhecimento, quais sejam: Ciência da Computação e Linguística.

1.1 Definição do Problema de Pesquisa

O aprendizado de uma disciplina acadêmica exige o aprendizado de novos conceitos, fazendo com que o indivíduo relacione conceito, termo e uma simbologia. Observa-se que, na sala de aula, onde são apresentados aos alunos novos conceitos, grande parte da comunicação se dá por meio da língua oral. Para os alunos surdos, esse tipo de comunicação não é suficiente, o que torna ainda mais difícil a integração destes no ambiente acadêmico. A Figura 6 ilustra a apresentação de um conceito na sala de aula, o qual o professor apresenta por meio da língua

Figura 4 – A apresentação de conceitos na sala de aula tradicionais.



Fonte: Próprio autor.

oral, escrita e por meio de ilustrações conceituais, como no caso do conceito de "fachada". No entanto, essa abordagem não atinge o aluno surdo.

Para que o aluno surdo consiga se comunicar com outros alunos (ouvintes) e o professor, faz-se necessária a presença de um intérprete, que faça uma tradução do que está sendo apresentado aos alunos. Porém, a tradução não é direta, e em alguns casos não há nem sinal para o conceito apresentado. Isso torna necessária e, até mesmo comum, a criação de novos sinais pelo intérprete e pelo aluno surdo.

Figura 5 – Há grande possibilidade de criação de sinais diferentes para o mesmo conteúdo, em diferentes instituições de ensino, visto que frequentemente os trabalhos de criação de tais sinais são feitos de forma independente e sem considerar regras ou critérios específicos.

Fonte: Próprio autor.

Os sinais terminológicos em boa parte são criados pelo intérprete e pelo aluno surdo durante um novo conteúdo e são utilizados durante o aprendizado. A criação destes sinais não segue critérios ou regras pré-estabelecida. O que em geral ocorre é uma associação entre o conceito e o sinal criado, e o comum é que estes sinais sejam perdidos uma vez que não são armazenados ou escritos, ou então por, simplesmente, não serem mais utilizados pelos indivíduos. Além disso, a criação do sinal ocorre em vários ambientes, e nem sempre há intercâmbio dos sinais, o que acarreta na existência de sinais diferentes para o mesmo conceito. Esse problema é ilustrado pela Figura 5.

Considerando a importância do aprendizado por meio do relacionamento entre os termos e conceitos e a carência de terminologia em CTEM em língua de sinais, a presente tese propõe uma abordagem inovadora, baseada em métodos computacionais, para tratar este problema. Especificamente, propõe-se: (1) a extração de descritores de forma a partir de uma possível imagem contendo a ilustração do significado do conceito técnico sem sinal de interesse, utilizando-se visão computacional, e (2) a determinação do radical do termo associado ao conceito na língua oral de referência do surdo, utilizando-se processamento de linguagem natural. Ao comparar essas informações associadas ao conceito técnico sem sinal com as correspondentes de sinais de propósito geral em um repositório, a abordagem desenvolvida cria um novo sinal semanticamente aceitável a partir de sinais pré-existentes com características semelhantes.

Para tanto, este trabalho considera duas hipóteses principais:

- É possível capturar a informação de iconicidade associada ao conceito sem sinal, a partir de uma ilustração ou imagem representativa deste, por meio da utilização de algoritmos de visão computacional;
- É possível simular a arbitrariedade no processo de criação do sinal, a partir dos modelos de sinais propostos pela abordagem, baseada na alta similaridade entre o radical do termo na língua oral de referência associada ao conceito sem sinal e os radicais dos termos cujos sinais são utilizados como modelos.

Finalmente, assume-se neste trabalho, sem perda de generalidade, que para cada conceito sem sinal de interesse, é possível prover uma ilustração ou imagem representativa.

1.2 Motivação

A principal motivação desta tese está relacionada aos problemas e às dificuldades encontradas pela pessoa surda, durante o aprendizado escolar e sua participação ativa no mercado de trabalho.

Os setores-alvo priorizados para a aplicação dessa nova abordagem computacional são aqueles com maior relação direta com a língua de sinais: institutos para pessoas surdas que utilizam Libras como forma de comunicação e instituições de formação de intérpretes, que auxiliam na tradução e na transmissão português-Libras.

O aprendizado do surdo envolve a transposição de inúmeras barreiras, desde a falta de orientação à família e ao próprio surdo acerca da aceitação e dos processos educacionais e de cidadania, até a falta de recursos didáticos no aprendizado da língua de sinais. O fato de o vocabulário ser empobrecido é semelhante a um bloqueio do pensar, e ocasiona a geração de um desenvolvimento inadequado das habilidades dos indivíduos (GLICKMAN, 2007, p. 143).

Nesse contexto, associado à educação dos surdos, vários autores descrevem que há uma grande carência por materiais didáticos, principalmente no âmbito da Linguística.

A educação do surdo envolve tanto o aprendizado dos surdos, quanto a qualificação dos intérpretes e professores. Além disso, se faz necessária também a busca, seja em bibliotecas ou num domínio na internet, por novas metodologias e conteúdos, como novas terminologias específicas em língua de sinais, que permitam a sua aplicação em sala de aula.

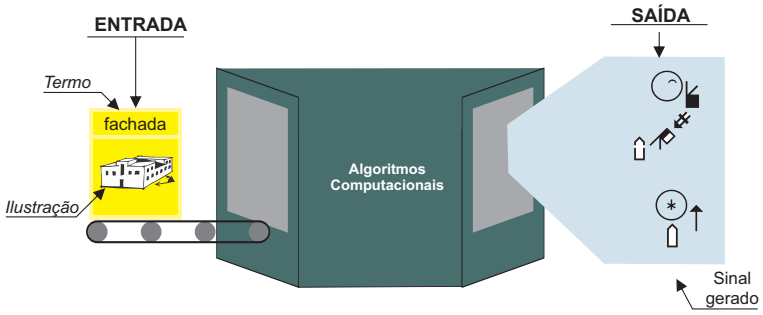
Não se pode deixar de relatar também o impacto social deste trabalho, uma vez que permitirá a acessibilidade das pessoas surdas em áreas nunca antes acessadas, sejam elas ONG's, órgãos do governo, projetos sociais, entre outras.

Stadler, Filietaz e Hussein (2012) mostram em seu trabalho que há uma grande necessidade de sinais para a disciplina de química, o que faz com que o aluno surdo, que esteja cursando essa disciplina, não tenha acesso a boa parte do conhecimento a ser desenvolvido. Suprir essa ausência de sinais específicos para os termos conceituais, seja da área de química, como foi apresentando também no trabalho de Stadler, Filietaz e Hussein, Fabiana (2014), ou das outras áreas em geral, é de suma importância para o desenvolvimento e para a dispersão do conhecimento por meio dos sinais elaborados, consolidando, assim, uma grande possibilidade de melhoria do ensino para os sujeitos da comunidade surda. Neste cenário, a presente tese abordará em particular a escassez do Léxico Terminológico da língua Libras (NASCIMENTO, 2009; FELIPE, 2006; STUMPF, 2005), utilizada como estudo de caso para avaliar a abordagem proposta.

1.3 Objetivos: Geral e Específicos

O objetivo geral desta tese consiste em desenvolver e em avaliar uma nova abordagem computacional que permitirá a geração de novos sinais técnicos para a língua de sinais. Para tanto, a abordagem partirá de uma entrada contendo uma ilustração representativa do conceito técnico sem sinal e uma palavra de uma dada língua oral, utilizará de algoritmos da

Figura 6 – Uma visão da abordagem proposta para geração de novos sinais técnicos para língua de sinais.



Fonte: Próprio autor.

computação para gerar automaticamente sinais específicos e inéditos para áreas técnicas, cujo desenvolvimento atualmente é realizado somente por humanos. Isso é bastante dispendioso, conforme é possível observar, nos trabalhos de Sousa e Silveira (2010), Stadler, Filietaz e Hussein (2012) e Lima (2014).

Para que o objetivo geral deste trabalho fosse devidamente atingido, os seguintes objetivos específicos foram considerados:

- Implementar um algoritmo de visão computacional que permita descrever a forma de uma imagem ou ilustração associando a um termo sem sinal, buscando assim o teoria iconicidade em relação à forma;
- Implementar um algoritmo de processamento de linguagem natural, que permita extrair o radical do termo da língua oral, para auxiliar na criação do sinal;
- Definir a(s) área(s) de conhecimento em CTEM, para as quais são criados neologismos terminológicos e construir um repositório de dados, compostas por: (i) listas de termos em português referentes a conceitos sem sinal em Libras; (ii) ilustrações técnicas de tais conceitos; (iii) sinais de propósito geral em Libras, acompanhados de seus correspondentes termos em português e ilustrações técnicas. Estes repositórios de dados são essenciais para avaliação da abordagem, por conterem os insumos utilizados pela mesma para gerar novos sinais;
- Divulgar cientificamente, após a devida proteção intelectual, os resultados obtidos nesta tese por meio da publicação de artigo em periódico internacional e trabalhos em eventos nacionais e/ou internacionais.

1.4 Contribuições

Uma vez que o desenvolvimento ou a geração de novos sinais é algo extremamente necessário e que demanda muito tempo para a sua criação, a principal contribuição deste trabalho é uma abordagem inovadora para auxiliar a produção de dicionários em CTEM, de forma parametrizada, dinâmica e escalável, com base na geração de novos sinais por meio de métodos computacionais dos campos de visão computacional e do processamento de linguagem natural.

Ao contrário dos esforços existentes no desenvolvimento de glossários lexicais para língua de sinais, que em geral são baseados em comitês de especialistas ou mesmo em trabalhos colaborativos, envolvendo diferentes pessoas da comunidade surda, a solução proposta é capaz de criar, com o apoio de uma metodologia assistida por computador, uma quantidade significativa de sinais de qualidade razoável a uma velocidade mais alta do que os métodos tradicionais, e com base em critérios transparentes e objetivos.

Permite-se, assim, uma acessibilidade aos surdos, de maneira geral, a áreas a que poucos têm acesso, sendo esta a principal contribuição social deste projeto. Além da característica social, existe o lado científico e tecnológico, dado o fato de criar um conjunto de informações dos sinais até então inexistentes, permitindo, assim, que estudiosos da língua de sinais possam utilizar este conteúdo transformador.

Cabe ressaltar que não é o objetivo desta abordagem propor uma solução que substitua a criação de novos sinais por meio dos seres humanos, mas sim auxiliá-los por meio de uma ferramenta computacional. Finalmente, deve-se ressaltar a não identificação de trabalhos na literatura cujas soluções para criação de sinais sejam apoiadas por métodos computacionais, fato este que sugere que a abordagem proposta nesta tese é potencialmente a primeira desta natureza.

1.5 Organização do trabalho

O presente capítulo apresenta uma contextualização geral, o problema de pesquisa, além de descrever as principais motivações e os objetivos que esta tese pretende alcançar. No Capítulo 2, são apresentados os principais conceitos e as técnicas fundamentais para o desenvolvimento da abordagem computacional proposta. O Capítulo 3 contextualiza e discute os trabalhos relacionados que abordam, em certa medida, o problema de pesquisa desta tese, descrevendo, assim, o estado da arte. A abordagem proposta por esta pesquisa é apresentada no Capítulo 4, em que são explanados os principais métodos e técnicas desenvolvidos com o objetivo de solucionar o problema de pesquisa. No Capítulo 5 são expostos os resultados experimentais obtidos com a finalidade de validar a presente tese. Por fim, no Capítulo 6 são apresentadas as conclusões, bem como a indicação de trabalhos futuros a serem desenvolvidos.

Capítulo 2

Fundamentação Teórica

Neste capítulo são apresentados os principais conceitos e técnicas que fundamentam a presente tese e que contribuíram para o desenvolvimento de metodologias para a criação de novos termos em línguas de sinais, em específico, o desenvolvimento de sinais-termos[1] em Libras.

Para tanto, na seção 2.1 será feita uma descrição detalhada sobre as características linguísticas das línguas de sinais, bem como sobre os princípios norteadores que descrevem o processo de formação do sinal. Em seguida na seção 2.2, são brevemente descritos alguns desafios relacionados à educação de pessoas surdas, bem como suas interações com as novas tecnologias.

Por fim, são abordados nas seções 2.3 e 2.4, conceitos básicos e técnicas das áreas de Visão Computacional e Processamento de Linguagem Natural, os quais fundamentam a abordagem computacional proposta para apoiar a criação de novos termos em línguas de sinais.

2.1 Línguas de Sinais

Cada país possui sua própria língua oral, bem como sua própria língua de sinais, sendo tais línguas completamente independentes. Por exemplo, há a Língua Brasileira de Sinais (Libras) no Brasil, a Língua Gestual Portuguesa (LGP) em Portugal, a Língua Britânica de Sinais (BSL) no Reino Unido e a Língua Americana de Sinais (ASL) nos Estados Unidos. Não há uma relação direta entre a língua oral e o seu referente da língua de sinais, além disso não há uma relação entre conceitos visuais do seu referente, ou seja, as línguas orais não são onomatopaicas [2], logo as línguas de sinais não são meramente mímicas (JOHNSTON, 1989; STUMPF, 2005).

[1] "Sinal-termo. 1. Termo da Língua de Sinais Brasileira que representa conceitos com características de linguagem especializada, próprias de classe de objetos, de relações ou de entidades. 2. Termo criado para, na Língua de Sinais Brasileira, denotar conceitos contidos nas palavras simples, compostas, símbolos ou fórmulas, usados nas áreas especializadas do conhecimento e do saber. 3. Termo adaptado do português para representar conceitos por meio de palavras simples, compostas, símbolos ou fórmulas, usados nas áreas especializadas do conhecimento da Língua de Sinais Brasileira."(FAULSTICH, 2014, p.1)

[2] Onomatopaicas ou onomatopeia: figura de língua que reproduz um som com um fonema ou palavra.

A comunicação por meio de gestos é abordada, por exemplo, pelo trabalho de Rizzolatti e Arbib (1998), os quais avaliam a proposta de um sistema de correspondência de observação / execução, visando-se fornecer uma ponte necessária entre o "fazer" e o "comunicar".

As línguas de sinais, como as línguas orais, possuem características específicas, possuindo nove características linguísticas que foram propostas por Hockett e Hockett (1960), sendo elas: (1) o modo de comunicação; (2) semanticidade; (3) função pragmática; (4) permutabilidade; (5) transmissão cultural; (6) arbitrariedade; (7) diferenciação; (8) deslocamento; e (9) produtividade.

Além dessas características possuem gramática própria e podem ser utilizadas e analisadas, como qualquer outra língua, permitindo discutir quaisquer tipos de assunto seja do mais simples ao mais complexo e até mesmos os assuntos abstratos.

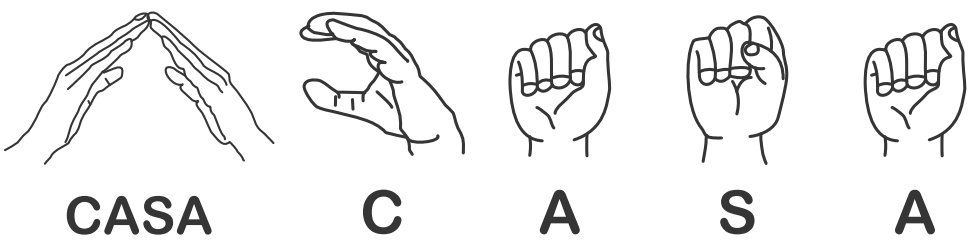
O falante da língua de sinais em sua comunicação se utiliza da apresentação do sinal por meio de movimento corporal para apresentá-lo, o que difere da língua oral que se utiliza do sistema respiratório para reprodução do som. Na língua de sinais quando não há um sinal para um terminado conceito técnico e para nomes próprios é utilizado a datilologia, da palavra escrita em língua oral. Seria semelhante a soletrar a palavra. Para ilustrar, a palavra "CASA", em Libras é apresentado pela Figura 7a e a datilologia do mesmo termo é apresentado pela Figura 7b. É possível observar a junção das letras em português ao formar o termo, porém não ocorrem na Libras. A datilologia quando utilizada em algumas vezes pode ser fonte de novos sinais, chamados de sinais lexicalizados (PARTON, 2006; FELIPE, 2007; NAKAMURA, 1995).

2.1.1 As relações entre Línguas de Sinais e Línguas Orais

A língua de sinais em sua essência linguística possui gramática própria sendo diferente da língua oral. Além disso, língua de sinais utilizadas em países diferentes possuem gramáticas próprias como, por exemplo, no caso da ASL e da BSL, em que, mesmo sendo o inglês a língua oral utilizada nos Estados Unidos e na Inglaterra, tais línguas de sinais são diferentes entre si, embora reciprocamente inteligíveis (BIBER et al., 1999; EMMOREY, 2001; HUDDLESTON;

Figura 7 – Ilustração do termo "CASA": 7a em Libras e respectivamente em português; 7b datilologia do termo e as letras alfabéticas.

(a) O sinal e a palavra. (b) A datilologia e as letras



Fonte: Próprio autor.

PULLUM, 2002). É interessante ressaltar que a gramática da ASL possui uma maior relação de similaridade com língua oral japonesa do que com à língua inglesa. Ademais, os números de línguas orais e de sinais não possuem relação, ou seja, não são os mesmos. É o caso, por exemplo, da África do Sul, que possui uma grande quantidade de línguas orais, oficiais e não oficiais, amplamente utilizadas, e apenas duas línguas de sinais (NAKAMURA, 1995; AARONS; AKACH, 1998; LIMA, 2014).

As velocidades de pronunciamento das sentenças das línguas orais e de sinais são semelhantes conforme apresentado no estudo de Bellugi e Fischer (1972). Nesse estudo, foi solicitado que os participantes contassem alguma história que conheciam plenamente e que apresentassem, em ordens diferentes, três interpretações da história. No primeiro momento, cada história, foi contata em ASL, posteriormente na língua oral inglesa e por fim em ASL e inglesa simultaneamente. Esse processo todo foi registrado por meio de filmagens e de transcrições, tendo como resultado uma média de 4,7 palavras por segundo e 2,3 sinais por segundo. Não obstante a utilização de mais palavras do que sinais, foi necessária uma média de 122 sinais por história; por outro lado, na língua oral foram necessárias 210 palavras.

O Quadro 1, fundamentado no trabalho de Quadros e Karnopp (2004), tem como objetivo apresentar algumas das principais diferenças entre a língua portuguesa e a Libras a fim de subsidiar o desenvolvimento desta tese e não tem como objetivo discutir pontos de vistas linguísticos.

Quadro 1 – Diferenças entre produções na língua portuguesa e em Libras.

Característica	**Libras**	**Língua Portuguesa**
Modalidade	Visual-espacial	Oral-auditiva
Base em experiências	Visuais da comunidade surda	Auditivas do ouvintes
Sintaxe	Espacial, incluindo os classificadores	Linear, utilizando a descrição para captar o uso de classificadores
Estrutura tópico-comentário	Utiliza esta estrutura	Evita esta estrutura
Estrutura de foco através de repetições sistemáticas	Utiliza esta estrutura	Não é comum o uso desta estrutura
Referências anafóricas[3]	Pontos estabelecidos no espaço	Pronomes
Marcação de Gênero	Não possui	Possui, a ponto de ser redundante
Expressões faciais	Atribui valor gramatical	Não é considerado relevante, apesar de poder ser substituído pela prosódia
Escrita	Não é alfabética	É alfabética

Fonte: Quadros e Karnopp (2004)

[3]Anáfora é um procedimento sintático que consiste em repetir um elemento anteriormente expresso.

Além disso, tem-se que as línguas orais são lineares, isto é, um único som é produzido ou recebido por vez, já as línguas de sinais utilizam atributos do meio visual espacial e toda uma cena pode pertencer a mensagem. Estes atributos ou características espaciais do meio visual é analisado na língua de sinais por classificadores, dado que por eles é possível reconhecer tamanho, tipo, forma, movimento e extensão (TOLAR et al., 2008; PIVETTA; SAITO; FLOR, 2013; LIMA, 2014).

2.1.2 Sinais Manuais e Não Manuais

No sinal, frequentemente há presença da mão, no entanto, o corpo, a cabeça as sobrancelhas, os olhos, o nariz, as bochechas e a boca podem também compor o sinal, sendo movimentados e combinados para mostrar informações estruturais como, por exemplo, distinção lexical, conteúdo adjetival ou adverbial e funções discursivas. Nas línguas orais este tipo de estrutura é denominado prosódia, a qual é apresentada por meio dos acentos e da entonação (FELIPE, 2013; EMMOREY et al., 2014).

A utilização dos sinais não manuais foi avaliada no trabalho de Arrotéia (2005), como, por exemplo, a negação em sentenças, que em geral é composta por movimentos faciais que são de sobrancelhas abaixadas, o contorno da boca é alterado ou a cabeça é girada repetidamente no eixo horizontal.

Para exemplificar a ideia de um sinal não manual, temos como, por exemplos, os termos "LADRÃO" e "COITO" apresentados pelas Figuras 8a e 8b. Observa-se que no primeiro termo há o movimento bucal, da bochecha e da língua fazendo um movimento do lado esquerdo da boca para o meio. Entretanto, o segundo termo é apresentado por um sinal composto pelo movimento bocal e o movimento da bochecha.

Figura 8 – Sinais não manuais dos termos "LADRÃO": 8a em Libras e respectivamente em português; 8b Libras e respectivamente em português.

(a) Termo LADRÃO.

(b) Termo COITO.

Fonte: (FELIPE, 2013, p. 78)

Exemplo: Fiz uma compra na loja do **Pedro**, paguei-**lhe** e **ele** agradeceu a preferência.

2.1.3 Parâmetros Fonéticos e Fonológicos das Línguas de Sinais

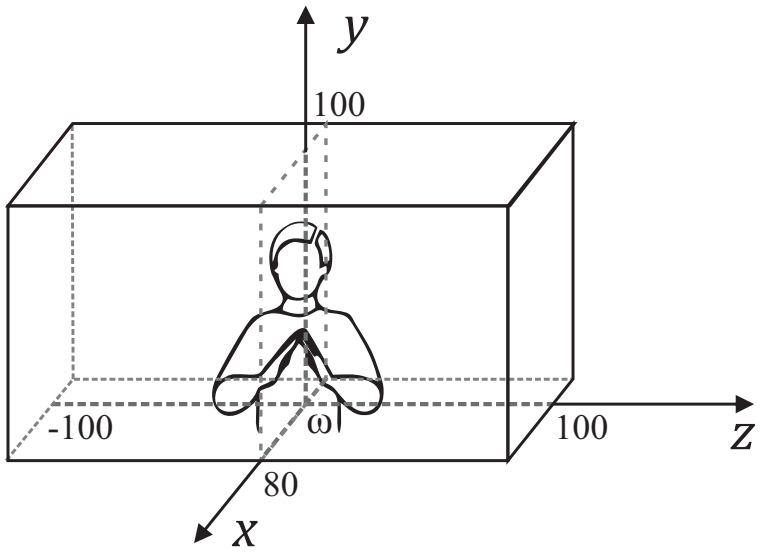
Como já mencionado anteriormente, as línguas de sinais em geral possuem em seu sinal a presença do movimento das mãos. Isto ocorre por se tratar de uma língua de modalidade gestual-visual/viso-espacial por suas informações linguísticas serem recebidas por meio dos olhos e transmitidas pelo movimento das mãos. Em uma análise da linguística sobre as palavras da língua oral são avaliados parâmetros fonéticos e fonológicos; isso também é avaliado acerca da língua de sinais, mesmo havendo diferenças entre as línguas de sinais e orais, seu emprego tem sido utilizado como referência em estudos aos elementos básicos das línguas de sinais.

A fonologia é responsável pela área que trata de identificar a estrutura e a organização dos constituintes fonológicos, propondo modelos descritivos e explanatórios. Logo, tem-se que a fonologia define quais são as unidades mínimas que formam o sinal e estabelece quais são os padrões possíveis de combinação entre essas unidades e as variações admissíveis (QUADROS; KARNOPP, 2004).

A origem da análise sobre a ótica da linguística teve como precursor o trabalho Stokoe (2005) em 1960, que depois foi reimpresso em 2005. Stokoe (2005) expôs uma estratégia linguística estrutural para analisar a formação dos sinais e propôs os termos "quirema" e "quirologia". O *quirema*, que seria semelhante ao *fonema* da língua oral, é o que define os três parâmetros básicos de um sinal, que são: (1) configuração de mão; (2) locação; e (3) movimento. Quanto ao termo *quirologia*, que seria semelhante ao *fonologia* da língua oral, tem-se a função de determinar as regras de combinação desses parâmetros. No entanto, com desenvolvimento de novas pesquisas, esses termos não foram incorporados. Vale ressaltar, por exemplo, que nos trabalhos de Battison (1974) e Battison (1978) afirma-se que o emprego dos termos *fonema* e *fonologia* para as línguas sinais é correto, uma vez que se baseiam na teoria linguística e são independentes do seu meio de comunicação. Portanto, os parâmetros que podem compor um sinal são:

- Configuração de mãos: está relacionada às diversas posições que as mãos tomam na realização do sinal, podendo ser do alfabeto manual ou de outras formas. Durante a apresentação de um sinal, a configuração de mão não necessariamente permanece a mesma;
- Locação: condiz com a área do corpo, ou com o espaço dentro de um raio de alcance das mãos em que os sinais são apresentados. A Figura 9 ilustra o espaço de locação;
- Movimento: movimentos realizados com a mão, com o pulso e com os antebraços;
- Orientação da mão: direção para a qual a palma da mão aponta na produção do sinal;
- Expressão facial e/ou corporal: expressões não manuais (movimento da face, dos olhos, da cabeça ou do tronco).

Figura 9 – Locação do sinal - espaço do sinal.



Fonte: Adaptado de Brito e Langevin (1994, p. 20).

Os fonemas são parte integrante dos morfemas. Assim como nas línguas orais esses também ocorrem nas línguas de sinais. Logo, a diferença está no fato de que nas línguas de sinais os fonemas podem ser articulados simultaneamente, no entanto, nas línguas orais isso não é possível, pois os fonemas possuem uma ordem linear que nada mais é que a sequência horizontal no tempo.

2.1.3.1 Restrições na Formação de Sinais

A formação de um sinal nas línguas de sinais está sujeita a limitações físicas e linguísticas. Essas limitações se devem inicialmente ao próprio sistema de apresentação, ou seja, ao sistema visual e ao sistema articulatório, próprios das articulações fonológicas que a anatomia e fisiologia das mãos permitem. A face, a região em que o interlocutor fixa o olhar, é aquela na qual os sinais são melhor reconhecidos (HIBBARD; FELS, 2011; PERNISS; VIGLIOCCO, 2014). Outras áreas podem não apresentar uma discriminação visual pouco precisa, ficando dependente da visão periférica dos comunicadores (QUADROS; KARNOPP, 2004).

Battison (1974) propôs duas restrições de estrutura de morfemas, as quais sejam, a Condição de Simetria e a Condição de Dominância, uma vez que são utilizadas as duas mãos e que os sinais com duas mãos são simétricos. Com relação à Condição de Simetria, tem-se que, se ambas as mãos estão se movendo na produção do sinal, as especificações de forma e a orientação das mãos devem ser idênticas, sendo assim o movimento deve ser simultâneo ou alternado. Para Condição de Dominância, quanto aos sinais que possuem formas de mão não idênticas, uma mão deve permanecer estática, ou seja, ficar passiva, enquanto a outra pode ser definida como a mão dominante, que executa o movimento. Essas restrições limitam a complexidade dos sinais, tornando-os em certa medida, mais fáceis de produzir e de compreender.

2.1.3.2 Criação de Sinais

No que diz respeito ao estudo do processo de criação de sinais, especialmente em Libras, tem-se como principais trabalhos os de Quadros e Karnopp (2004), Felipe (2006), Nascimento (2009) e Lima (2014), sendo esses referências fundamentais para o desenvolvimento deste trabalho. São poucos os trabalhos que tratam diretamente a questão da formação de sinais, principalmente em relação a Libras. O mesmo acontece em relação aos mecanismos de avaliação dos sinais já existentes, ou seja, quanto a sua análise morfológica.

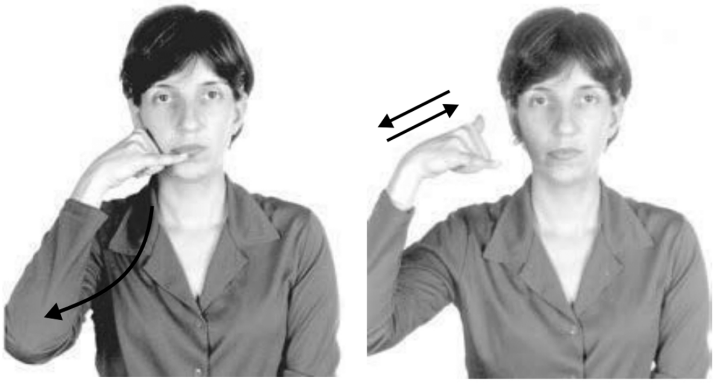
2.1.3.2.1 Morfologia

Para compreensão do processo de formação dos sinais, é necessário entender o aspecto da morfologia. Em Quadros e Karnopp (2004), os autores fazem uma ponderação da análise geral sobre a morfologia tradicional, apresentando duas áreas: a derivacional e a flexional. A derivacional relaciona diferentes palavras com a mesma estrutura lexical, por exemplo, a palavra em português "cantando" é derivada de "cantar". Já a flexional está relacionada à inclusão da informação gramatical à palavra que já existe, como flexões de número, tempo, gênero, aspecto e pessoa. Sobre os aspectos da língua de sinais, os autores relatam a complexidade da representação morfológica, e defendem que esta deveria ser abordada pela teoria lexical. É observado que o padrão de recombinação dos morfemas pelos sinalizadores decodificam a forma-base e a flexão morfológica separadamente.

Alguns processos de formação derivacional estão presentes na Libras assim como ocorre na língua portuguesa como, por exemplo, a derivação dos nomes dos verbos que se dá pela inclusão de um sufixo.

Na Libras, o processo de derivar nomes de verbos é ilustrado pela Figura 10, que mostra o

Figura 10 – Processo de derivação dos nomes dos verbos.



Fonte: Quadros e Karnopp (2004, p. 97)

Figura 11 – Processo de composição.

ESCOLA

Fonte: Quadros e Karnopp (2004, p. 103)

verbo "telefonar" e a palavra "telefone", destacando a mudança de movimento. Quanto ao processo por composição, que no português ocorre, por exemplo, com as palavras "ciclo" + "via" resultando em "ciclovia", é observado pelos autores que a Libras segue as mesmas regras da ASL, que são: regra do contato; regra da sequência única e a regra da antecipação da mão não-dominante, conforme é ilustrado na Figura 11, em que a palavra "escola" é composta pelos sinais "casa" + "estudar". Quadros e Karnopp (2004) relatam a existência de limites entre a fonologia e a morfologia, ainda sendo campos que devem ser pesquisados pois há uma carência de descritores e análises para a Libras.

A criação ou a formação de sinais é ainda um campo de pesquisa exploratório, ou seja, não há uma consolidação sobre a análise da influência da geração de sinais (QUADROS; KARNOPP, 2004; FELIPE, 2006; NASCIMENTO, 2009).

O processo de formação do sinal existe e, muitas vezes, é descrito à luz da língua oral. O falante de uma língua forma o sinal valendo-se da língua que utiliza todos os dias, obedecendo intuitivamente às regras de usos e de costumes e, principalmente, utiliza a estrutura dessa língua. Essa estrutura é adquirida, na maioria das vezes, na infância. Para Saussure, algumas questões devem ser consideradas a respeito da imutabilidade do signo[4] linguístico. A arbitrariedade do signo, a que se poderia atribuir uma possibilidade teórica de mudança ou de mutabilidade, na realidade, preserva a língua de qualquer tipo de modificação advinda da ação consciente de um falante da língua (NASCIMENTO, 2009; LIMA, 2014).

A criação dos sinais é de alta importância para o desenvolvimento dos termos específicos, mas também são importantes os mecanismos de registro para perpetuação da sua existência. Nesta seção foram descritos alguns fatores que auxiliam na análise do sinal e, consequentemente, em

[4] É a palavra que tem sentido, considerado signo linguístico.

sua geração. Várias pesquisas direcionam-se para dois aspectos principais que contribuem na formação do sinal, quais sejam o processo da arbitrariedade e o processo da iconicidade, os quais serão brevemente descritos na seção a seguir.

2.1.3.2.2 Noções sobre Arbitrariedade e Iconicidade

O processo de criação de sinais pode ser analisado como na língua oral, ao buscar subsídios para criação de novos termos. Como na análise do processo de formação de palavras, através da qual buscam-se elementos que permitam a associação entre conceito e seu significado. Nesse sentido, uma das primeiras obras a discutir estes aspectos é de 1970, sob autoria de Saussure (SAUSSURE, 1970), que em certa medida faz uma estruturação quanto ao estudos da linguagem.

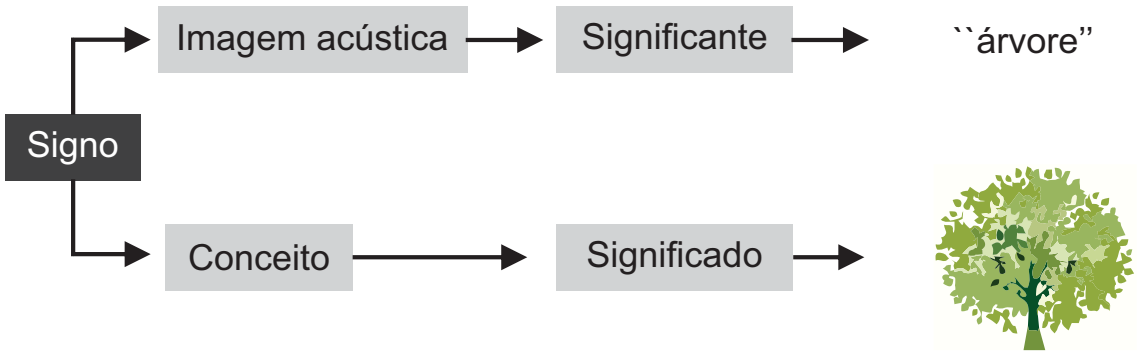
Alguns consideram que a arbitrariedade é o signo linguístico e é compreendida como a ausência de ligação entre a forma de uma palavra e seu significado (NASCIMENTO, 2009).

Para Saussure (1970), toda a linguagem é um sistema de signos, e o autor definiu que o signo une não uma coisa a um nome, mas sim o conceito de algo, como uma imagem acústica, a algum signo. No que diz respeito à imagem acústica não necessariamente é relacionada a um evento físico, mas principalmente à impressão que pode ser de forma, pegada, marca, ou seja, a relação que o cérebro faz à associação.

Para ilustrar essa definição, pode-se pensar no termo "árvore" sendo pronunciado entre dois falantes da língua oral. O som que é reproduzido por um sujeito transmissor é recebido pelo outro sujeito (receptor), que irá obter a pronúncia. De certa maneira, o sujeito ouvinte irá construir toda uma lógica que permitirá reconhecer a mensagem em formada em sua mente, ou seja, o signo transmite o conceito que é formado pelo seu significado e por fim a ideia formada, conforme é apresentado na Figura 12.

A relação arbitrária entre o significado e o significante pode ser algo motivado, ou seja, o que

Figura 12 – Diagrama que ilustra a proposta de Saussure.



Fonte: Adaptado de Fernandes (2011, p. 170).

para o princípio da iconicidade prevê motivação na relação entre forma e significado. Logo a iconicidade consiste na existência da compreensão entre a forma e o seu significado, sendo considerada como um fenômeno de cognição.

Sendo assim, a iconicidade permite associar um ícone ou um signo icônico a um conceito de algo real e, por isso, tem grande influência na formação de sinais em linguagens viso-espaciais. Para Simões (2009), a iconicidade, no âmbito mais genérico, não trata somente das questões voltadas à Língua de Sinais. O autor destaca que há uma forte influência dos ícones no processo de associação entre conceitos e imagem, demonstrando em vários aspectos a influência e o poder da iconidacide.

Nas língua de sinais, pode-se destacar que os sinais icônicos apresentam certa semelhança com uma visão sobre o mundo, como, por exemplo: o sinal de casa lembra a imagem de um telhado de uma casa, e o sinal de árvore é a representação do tronco e dos galhos de uma árvore. Em relação à corporificação e à iconicidade em Libras, pode-se dizer que existe relação icônica na utilização de partes do corpo para demonstrar a imagem real, como, por exemplo, nos sinais de pensar e de fome (NUNES, 2009).

A presença da formação de signos por meio imagético não é exclusividade da língua de sinais. Pode-se considerar uma característica relativa e não absoluta, várias línguas estão repletas de elementos icônicos. O que se questiona frequentemente em relação à iconicidade não é se uma forma é icônica, mas como ela é icônica (WESCOTT, 1971).

A motivação da construção do sinal por meio da iconicidade, com o passar do tempo, pode ter uma menor influência. E em Lima (2014) ressalta-se que, mesmo alguns sinais sendo de natureza icônica, existe a tendência para a mudança de direção dos sinais para a arbitrariedade, ao invés de se manter a iconicidade.

Nas línguas orais, a relação de iconicidade não é trivial, e alguns casos é percebida em termos que são derivados de onomatopeia. Em contrapartida, nas línguas de sinais, utiliza-se comumente a iconicidade para representar conceitos de objetos, movimentos e relações espaciais, a probabilidade de se utilizar somente a iconicidade para ideias abstratas é muito pequena. Sendo assim, iconicidade não é o único recurso para a criação de um sinal. Em Deuchar (2013), o autor propõe dividir os sinais em três categorias de formação: sinais arbitrários, sinais indéxicos e sinais icônicos.

2.1.4 Escrita de Sinais

Vários sistemas para escrita de línguas de sinais foram propostos. Um dos mais relevantes foi desenvolvido por Stokoe (1919-2000), com sua equipe da Universidade Gallaudet, os quais estabeleceram a ideia do caráter linguístico para a língua de sinais, criando também um

Figura 13 – Configurações de mãos propostas por Stokoe.

Fonte: Stumpf (2005, p. 48)

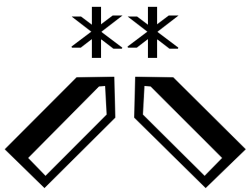
conjunto com cinco elementos: local (12 posições), as configurações de mãos (10 configurações), movimento que indica a ação (22 símbolos), orientação (4 indicações) e, por fim, os sinais diacríticos (duas possibilidades). A Figura 13 exibe a proposta para configuração das mãos.

O sistema proposto por Stokoe influenciou vários pesquisadores como Fraçois Xavier Neve, da Universidade de Liége, Bélgica. O sistema de notação foi publicado em 1996 com o objetivo de possibilitar uma enumeração e uma elaboração de sinais para informática, tendo como característica a escrita feita na vertical e de cima para baixo. Em geral, o sinal é monomanual, escrito em apenas uma coluna. Os signos, que são as menores partes de um sinal, são escritos na seguinte ordem: Configuração, Localização, Orientação e Ação.

Na Universidade de Hamburgo, na Alemanha, os pesquisadores Prillwitz, Vollhaber e seus colaboradores apresentaram o Hamnosys, um sistema de transcrição fonética. Ele tem em torno de 200 signos e teve o objetivo de ser internacional, ou seja, para qualquer Língua de Sinais. Utilizou-se de símbolos icônicos para melhor aprendizagem entre os usuários da língua, com número reduzido em relação à quantidade de símbolos. Esse sistema permitiu a integração com as ferramentas da computação pois apresentou uma sintaxe formal.

Por fim, tem-se o sistema de escrita proposto por Valerie Sutton, que dirige o *Deaf Action Commitee* (DAC), denominado *Sign Writing* (SW). A proposta inicial tinha como objetivo realizar notação de movimentos para dança e passou a ser desenvolvido e utilizado para língua de sinais, sem estar relacionado com uma língua específica, sendo propositadamente universal. É um sistema de representação gráfica dos signos das línguas de sinais, o qual permite por

Figura 14 – Ilustração da escrita em SW da palavra **casa**.



Fonte: Próprio autor.

meio desses representar as configurações das mãos, os movimentos, as expressões faciais e os deslocamentos corporais. Para ilustrar uma representação da escrita por meio do SW, a palavra "casa" é apresenta na Figura 14.

2.1.4.1 *Sign Writing*

O sistema de escrita de sinais denominado *Sign Writing* (SW) possui uma estrutura composta por 7 categorias (Mãos, Movimento, Dinâmica, Cabeça e Faces, Corpo, Localização detalhada e Pontuação), sendo que essas categorias podem possuir, por sua vez, subdivisões, como, por exemplo, no caso das mãos, que possui 10 grupos.

A estrutura computacional para representação destes símbolos é desenvolvida e disponibilizada no site do *Sign Writing* [5] que tem como seu fundadora a Sutton (1994), na qual desenvolveu a proposta do sistema de escrita de sinais *Sign Writing*. A evolução do mecanismo para armazenamento da escrita SW foi inicialmente desenvolvida por Rich Gleaves, que propôs um editor de símbolos associado a um *FileMaker* (um gerenciador simples de dados), que se relacionava a um identificador (ID). Por exemplo, a mão é definida pelo código **001-01-01-01**. Esta codificação foi denominada *International SignWriting Alphabet* (ISWA) e posteriormente utilizou-se da estrutura do XML. Para exemplificar, a palavra "casa" descrita em XML é apresentada pelo código abaixo, sendo que são descritos os símbolos utilizados, posteriormente sua posição na sequencia de sua.

```
<signbox>
  <seq>S15a11</seq>
  <seq>S15a19</seq>
  <seq>S20600</seq>
  <sym width="23" height="23" left="-25" top="-4">S15a11</sym>
  <sym width="23" height="23" left="2" top="-4">S15a19</sym>
  <sym width="22" height="11" left="-11" top="-19">S20600</sym>
</signbox>
```

[5] <=> http://www.signwriting.org

Em Costa e Dimuro (2003), os autores propõem uma linguagem baseada no XML denominada SWML. Esta linguagem utiliza-se dos conceitos de marcação decorrente do XML, sendo altamente descritiva em relação ao sinal, conforme é exemplificado na Figura 15.

Atualmente o SW é codificado por meio da linguagem *Formal and Regular SignWriting* (FSW) (SUTTON, 2006; KOLLER; NEY; BOWDEN, 2013). O mecanismo de representação por meio do FSW é uma linguagem formal de estrutura regular, que codifica de forma compacta os signos. A representação em SW da palavra casa apresentada na Figura 14 tem a seguinte representação em FSW:

`AS15a19S15a11S20500M525x519S15a19475x496S15a11502x496S20600489x481`

Para identificação entre os signos e o ID, são usados os caracteres em ASCII, denominados símbolos-chave (SLEVINSKI, 2013). Os símbolos-chave são expressos por um conjunto composto de 6 caracteres, tornando a chave e seus símbolos explícitos enquanto, por outro lado, é mantida

Figura 15 – Exemplo da linguagem SWML.

```
<signbox>
 <symb x="45" y="59" x-flop="0" y-flop="1" color="0,0,0">
   <category>01</category>
   <group>05</group>
   <symbnum>001</symbnum>
   <variation>01</variation>
   <fill>02</fill>
   <rotation>01</rotation>
 </symb>
 <symb x="79" y="33" x-flop="1" y-flop="1" color="0,0,0">
   <category>02</category>
   <group>10</group>
   <symbnum>003</symbnum>
   <variation>01</variation>
   <fill>01</fill>
   <rotation>01</rotation>
 </symb>
 <symb x="46" y="19" x-flop="0" y-flop="1" color="0,0,0">
   <category>02</category>
   <group>10</group>
   <symbnum>005</symbnum>
   <variation>01</variation>
   <fill>02</fill>
   <rotation>01</rotation>
 </symb>
 <symb x="67" y="66" x-flop="0" y-flop="0" color="0,0,0">
   <category>01</category>
   <group>05</group>
   <symbnum>001</symbnum>
   <variation>01</variation>
   <fill>02</fill>
   <rotation>01</rotation>
 </symb>
</signbox>
```

Fonte: Costa et al. (2004, p. 4)

Tabela 1 – Definição das marcações em SW.

Marcação	Descrição
A	Sequência de marcação inicial
B	Marca do SingBox
L	Esquerda em relação ao centro
M	Ao centro
R	Direita em relação ao centro

Fonte: Adaptado do trabalho Slevinski (2013).

uma codificação implícita. O primeiro caractere da chave contém a letra "S", permitindo assim a análise e a avaliação humana da codificação. Os cinco caracteres restantes representam a codificação do símbolo em hexadecimal na base 16. Então, tem-se que a codificação do símbolo contém *número do símbolo base* = *"n"*, o *preenchimento* = *"f"*, e *rotação* = *"r"*, uma função denominada *dechex* que permite converter o número de decimal para hexadecimal, tal que:

símbolo-chave = 'S'+'dechex(n+255)'+'dechex(f-1)'+'dechex(r-1)'.

Além dos símbolos-chave, é necessário compreender outras marcações existentes no FSW, como o posicionamento dos símbolos, que estão definidos na Tabela 1.

O FSW é descrito como um roteiro em que há símbolos sequenciais definidos em posições bidimensionais, semelhante à ideia de um layout para os símbolos. Sendo assim, existe um agrupamento de símbolos para representar um sinal, sendo dispostos verticalmente de cima para baixo e alinhados ao centro. Podem ocorrer também sinais que são dispostos na horizontal da esquerda para a direita. Os símbolos têm tamanhos estáticos, ou seja, com largura e com altura definidas. O SingBox é a descrição de como ocorre essa distribuição dos símbolos no plano bidimensional, portanto uma representação em cadeia normativa, conforme demonstrado pela Figura 16.

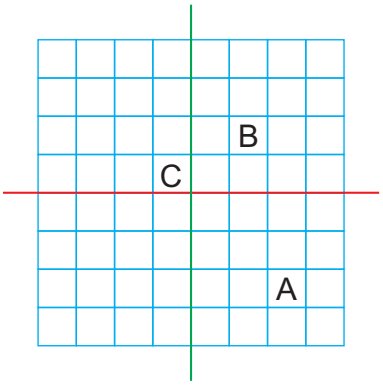
Na sequência não há uma premissa, ou seja, a ordem não é necessariamente ABC, ACB, BAC, BCA, CAB, ou CBA. O que existe é uma representação precisa, na qual o arranjo deve especificar a posição em coordenadas no plano bidimensional de cada elemento, como por exemplo: C (-2, -1), B (3, -2), A (5,2). Cabe ressaltar que a ordem sequencial dos elementos não é significativa, exceto em casos de sobreposição dos símbolos. As dimensões aceitas nos eixos X e Y estão na faixa de -250 à 249. No caso da codificação FSW, as coordenadas regulares são definidas por 7 caracteres ASCII, sendo: 3 dígitos seguidos pela letra 'X', seguidos por mais 3 dígitos. Os números variam de 250 a 749, sendo que 500 é o ponto central. Pode-se inferir que a sequência de caracteres "250" é igual ao valor numérico -250 e que "749" é igual ao valor 249. A validação dos dígitos é realizada pela expressão regular $[0-9]3 \times [0-9]3$, e a expressão regular para a faixa de 250 a 749 é dada por:

(2[5-9][0-9]|[3-6][0-9]2|7[0-4][0-9])x(249|2[5-9][0-9]|[3-6][0-9]2|7[0-4][0-9])

A modelagem da expressão regular do FSW é apresentada na Tabela 2. Os símbolos são expostos pelas sete categorias descritas a seguir:

1. **Mãos**: são organizados em 10 grupos, com base nos números 1 à 10 da ASL;
2. **Movimento**: são os símbolos de contato, ou pequenos movimentos dos dedos, definidos por setas retas, por setas com curvas e por círculos. Esses símbolos são agrupados em 10 grupos com base em planos: frontal, plana, incluindo o movimento que é paralelo à parede de frente e o plano de piso, que inclui o movimento que é paralelo ao piso.
3. **Dinâmica**: são símbolos usados para dar a ideia demonstrar sentimento ou de tempo em relação ao movimento. Fornecem ênfase ao movimento ou à expressão, que, quando combinados com a pontuação dos símbolos, tornam-se o equivalentes ao ponto de exclamação. Símbolos de tempo são usados para mostrar alternância ou movimento simultâneo.
4. **Cabeça e rosto**: são os símbolos que começam com a cabeça e, em seguida, a partir do topo do rosto com deslocamento para baixo;
5. **Corpo**: são considerados os movimentos do torso, ombros, quadris e as pernas, que são usadas nas Línguas de Sinais, e que fazem parte de sua gramática;
6. **Localização detalhada**: seu uso não é comum, mas detalha os pontos de articulação em relação aos signos. São usados na ortografia por meio da sequência de sinais, podendo ser uteis quando há grande quantidade de escritas, como em dicionários, em animação, ou no processo de tradução por meio de *scripts* e, também para a análise detalhada usada em pesquisas na área da linguística.
7. **Pontuação**: símbolos que são usados ao escrever sentenças ou documentos completos em *SignWriting*.

Figura 16 – SingBox - Layout de posicionamento dos símbolos SW.



Fonte: Adaptado do trabalho Slevinski (2013).

A partir do FSW é possível ter uma escrita compacta dos sinais e de alta relevância, por ser tratar de uma linguagem de sinais universal, ainda que tenha traços da ASL.

Além da preservação e da divulgação do sinal, o sistema proposto neste projeto pretende criar automaticamento novos sinais. Para isso irá se basear, entre outras coisas, nas formas de objetos existentes nas representações visuais dos termos inexistentes no léxico da língua de sinais considerada, o que de certa forma relaciona-se à propriedade da iconicidade, conforme descrito na seção a seguir.

2.1.5 Língua Brasileira de Sinais (Libras)

A Língua Brasileira de Sinais (Libras), pode-se dizer que teve início por volta do ano de 1855, quando o francês Eduardo Huet recebeu o convite de D. Pedro II para trabalhar com a educação para surdos. Logo, em 26 de setembro de 1857, na cidade do Rio de Janeiro, então capital do Império, foi criado o Imperial Instituto de Surdos-Mudos, que atualmente é chamado de Instituto Nacional de Educação de Surdos (INES) (CURRIE; MEIER; WALTERS, 2002). Alguns sinais utilizados hoje pelas comunidades surdas brasileiras foram criados por um ex-aluno do Instituto, Flausino José da Gama, que, em 1875, criou um pequeno vocabulário ilustrado de sinais brasileiros chamado *Iconografia dos Signaes dos Surdos-Mudos* (SOFIATO; REILY, 2012; LIMA, 2014)

Durante anos, a língua utilizada pelos surdos brasileiros não era reconhecida como tal. Somente em 2002, por meio de um plano linguístico que surgiu a partir da Lei Federal 10.436 (BRASIL, 2002), de abril de 2002, e do subsequente Decreto Federal 5.695 (BRASIL, 2005), de dezembro de 2005, que a Libras se consolidou no Brasil. Convém esclarecer que, anteriormente ao reconhecimento nacional, o Estado de Minas Gerais já havia reconhecido a Libras em 10 de janeiro de 1991 por meio da Lei 10.379 (ESTADO DE MINAS GERAIS, 1991). Tem-se que a partir desses movimentos de regulamentação foi assegurado aos surdos o direito de serem educados em sua própria língua e o direito de serem auxiliados por intérpretes em espaços públicos, educacionais e em meios de comunicações.

Tabela 2 – Definição da expressão regular do FSW.

Estrutura	Expressão regular
Símbolo-Chave	S[123][0-9a-f]2[0-5][0-9a-f]
Coordenada	[0-9]3x[0-9]3
Signbox	[BLMR]([0-9]3x[0-9]3)(S[123][0-9a-f]2[0-5][0-9a-f][0-9]3x[0-9]3)*
Sinal	(A(S[123][0-9a-f]2[0-5][0-9a-f])+) [BLMR]
	([0-9]3x[0-9]3) (S[123][0-9a-f]2[0-5][0-9a-f] [0-9]3x[0-9]3)*

Fonte: Adaptado do trabalho Slevinski (2013).

2.1.5.1 Dicionários em Libras

O desenvolvimento de um dicionário envolve questões multidisciplinares e fundamenta-se em diversos conceitos da linguística. Na linguística, a área que dá um certo suporte a este desenvolvimento é a terminologia, que tem o propósito de evitar a polissemia e as ambiguidades as quais são comuns na linguagem natural (KRIEGER; FINATTO, 2004).

O foco da terminologia é a unidade terminológica, ou seja, o termo, considerando os aspectos linguísticos, cognitivos e pragmáticos. Nesse sentido, para a construção de um dicionário, a terminologia permite a descrição detalhada dos termos. Logo, os termos constituem um subconjunto de signos linguísticos relativos a uma área conceitual e especializada, tendo papel fundamental no desenvolvimento de tesauros, glossários e em especifico os dicionários (PONTES, 1997).

No que diz respeito às línguas de sinais, Zwitserlood (2010) ressalta as dificuldades em desenvolver o dicionário para língua de sinais, seja sobre o aspecto de representação do sinal ou sobre as diferentes línguas de sinais existentes dentro do próprio pais.

Além disso, é notável a existência de uma grande lacuna no intervalo entre as produções lexicográficas da língua oral e da Libras, no caso do Brasil. Considerando a Libras, o primeiro trabalho como dicionário para língua foi o de Flausino José da Costa Gama, lançado em 1875, intitulado como "Iconographia dos Signaes dos Surdos - Mudos". Somente mais tarde, em 1969 o missionário americano Padre Eugênio Oates, que transitava pelo país, aprendeu a língua de sinais e, em consequência de suas viagens, observou que havia diferenças lexicais nos sinais dos surdos pelo país, lançando, assim, a segunda obra com a característica de dicionário, denominada "Linguagem das Mãos" (LIMA, 2014).

Posteriormente, em 2001, foi lançado o "Dicionário enciclopédico ilustrado trilíngue: Língua Brasileira de Sinais" de Capovilla e Raphael (2001a), sendo relançado posteriormente em 2008 como "Novo Deit - Libras - Dicionário Enciclopédico Ilustrado Trilíngue da Língua de Sinais Brasileira".

Além do dicionário apresentado anteriormente, outro dicionário de relevância na versão digital é o Dicionário da Língua Brasileira de Sinais (LIRA; SOUZA, 2011). Ele está em sua terceira versão disponível via internet e CD. Esses dicionários foram desenvolvidos com o propósito de conter os termos gerais da língua, ou seja, o que é usado no dia a dia das pessoas.

Embora a Libras tenha sido oficializada, seu léxico possui um número pequeno de termos; mesmo considerando que os dicionários apresentados acima são de propósito geral. O dicionário de Capovilla, por exemplo, em seus dois volumes possui 9.828 termos, enquanto o dicionário Houaiss da língua portuguesa possui mais de 442 mil entradas (HOUAISS; Mauro de Salles Villar, 2009). Considerando os dicionários terminológicos para as áreas de CTEM, esse cenário é

preocupante, e até a presente data, não há dicionários especializados em Libras, seja monolíngue ou bilíngue.

2.1.5.1.1 Dicionários Terminológicos em Libras

Até a presente data, a Libras não possui um dicionário terminológico ou até mesmo uma iniciativa para produção de obras lexicográficas terminológicas, seja formalmente ou oficialmente. O que se tem são trabalhos isolados que apontam para essa necessidade de haver terminologias específicas em Libras para as áreas de CTEM. Alguns trabalhos têm proposto a criação de mini-glossários, como, por exemplo, em Carmona (2015), a autora propõe um glossário para o ensino de Biologia, Sousa e Silveira (2010), por sua vez, propuseram um glossário para o aprendizado de Química e assim como Stadler, Filietaz e Hussein, Fabiana (2014).

Outras formas que buscam ampliar os termos específicos em Libras ocorrem em sala de aula, cabendo ao intérprete, ao surdo e ao professor essa árdua tarefa de produzir sinais para conceitos que são explicados em português. E de fato essa árdua tarefa tende a cair no esquecimento pois em geral não há registros destes sinais criados.

Cabe ressaltar o trabalho de Carmona (2015), que realizou um levantamento de quais termos são necessários no ensino de Biologia e, a partir disso, foi realizada uma pesquisa buscando os termos em língua de sinais, não só a Libras, mas outras línguas também para verificar se existiam sinais para os conceitos de Biologia. Como é comum, constatou-se que praticamente inexiste sinais para tais termos; sendo assim, foi formada uma equipe multidisciplinar para elaboração de um glossário para Biologia, para Ensino Médio, em Libras. Foram desenvolvidos nesse trabalho mais de 360 sinais para os conceitos identificados como necessários no aprendizado.

A cada dia vêm surgindo novas pesquisas que têm o objetivo de criar novos sinais específicos para as áreas de CTEM, porém é um trabalho que, além de árduo, não é escalável, e que em certa medida não possui qualquer tipo de fundamentação no processo de criação ou formação de sinais.

2.2 Educação de Surdos e Novas Tecnologias

A tecnologia vem permitindo uma maior comunicação entre as pessoas, antes limitada pela distância entre elas ou pelas diferenças entre as línguas. Considerando os surdos em específico, os dispositivos computacionais associados à Internet marcaram uma nova era, uma vez que as tecnologias atuais dão suporte visual, permitindo, assim, tanto que a comunicação dos surdos seja realizada por meio textual, quanto por imagens ou até mesmo por vídeos. E, em certa medida, o meio digital conflagrou a participação mais ativa dos ouvintes nas comunidades surdas, não obstante o mesmo aconteceu aos surdos (LU; HUENERFAUTH, 2011).

No processo educacional de estudantes surdos, os novos avanços tecnológicos têm propiciado mudanças importantes no processo de ensino-aprendizagem, sobretudo quando se considera que os alunos surdos necessitam que todo conteúdo que ocorre em seu aprendizado, seja em sala de aula ou até mesmo no laboratório, de forma falada ou sonora sejam apresentadas em formato visual, ou seja, em língua de sinais (WEST; SUTTON-SPENC, 2012). O que ocorre em geral é que há um intérprete de língua de sinais ou transcritores que fazem isso em tempo real. Quando ocorre o uso de elementos que são do estímulo visual para a transmissão do conhecimento, os estudantes que fazem uso da língua de sinal podem se beneficiar e tornar os conceitos-chaves mais claros (CAVENDER; LADNER, 2010).

É comum que os estudantes surdos utilizem várias fontes de informações visuais ao mesmo tempo, o que acarreta em perda de informações prejudicando os seus aprendizados (CAVENDER; BIGHAM; LADNER, 2009; KUSHALNAGAR, 2010). Pensando nesta problemática, o trabalho proposto por Kushalnagar, Cavender e Pâris (2010) propôs um sistema baseado em um arranjo de janelas, denominado Perspectivas de Exibição Múltipla (MVP). A proposta do MVP é auxiliar os alunos surdos, em sala de aula, por meio de janelas de vídeos, sendo que cada janela exibe uma visão de vídeo contextual de uma região específica. Comparados aos sistemas tradicionais, que têm objetivo de captura de palestras, a proposta habilita os alunos surdos a reunir de forma independente as regiões de interesse em uma única tela, superando as expectativas, o que a torna extremamente útil, uma vez que as salas de aulas contêm várias fontes visuais de informações, como demonstrações, slides, quadros.

Em Drigas et al. (2004), os autores propõem uma nova metodologia para auxiliar os estudantes surdos no ambiente virtual por meio da *e-Learning* adaptado para pessoas com deficiência auditiva. Nesse trabalho os autores utilizaram da teleconferência (sala de aula virtual) oferecendo um serviço de treinamento sob a forma de educação contínua para pessoas surdas, de forma fácil e amigável, com suporte de material eletrônico informativo adaptável.

É crescente o número de trabalhos que estão adotando as tecnologias para serem utilizadas como apoio no ensino de estudantes surdos. Em certa medida, o uso da tecnologia na educação oferece aos alunos e aos professores um método de apresentação de informações visuais, interativas e atrativas, permitindo a troca de pensamentos com todos os alunos, não somente os surdos.

A tecnologia na educação é um grande atrativo tanto aos estudantes quanto aos professores. Além da atratividade, vêm oferecendo métodos visuais de apresentação das informações, como se pode observar nos trabalhos de Kushalnagar, Cavender e Pâris (2010) e Drigas et al. (2004) este uso. Além disso, permite aos alunos, com qualquer grau de surdez, superarem as dificuldades no processo de aprendizado (ROBERSON, 2001).

2.3 Visão Computacional

Uma das características presentes no processo de formação de sinais está na associação do conceito a uma forma. Essa característica é descrita por vários autores e é denominada iconicidade (SIMÕES, 2009; GOMES; TEIXEIRA, 2014; DINIZ, 2010; NASCIMENTO, 2009; FELIPE, 2007).

Na área da computação existe uma subárea que trata dos mecanismos de reconhecimento de formas em informações visuais (por exemplo, imagens e vídeos), denominada Visão Computacional (VC). Há uma grande variedade de métodos que permitem detectar, descrever e reconhecer formas presentes nas imagens. Para este trabalho propõe-se a utilização do método denominado *Shape Context* (BELONGIE; MALIK; PUZICHA, 2000; BELONGIE; MALIK; PUZICHA, 2001), visto que este faz uso de um descritor que demarca pontos de interesse nas bordas detectadas de uma imagem com um pequeno número de pontos, comparados a outras metodologias, possuindo relativa tolerância a transformações na imagem, além de permitir uma recuperação eficiente de formas semelhantes. Essas propriedades contribuem com a proposta de recuperar formas semelhantes presentes nas imagens dos sinais já existentes, fazendo uma relação direta da iconicidade na imagem, por meio da forma.

2.3.1 Descritor de Forma

O método *Shape Context*, que foi proposto por Belongie, Malik e Puzicha (2000), em síntese, tem o objetivo de descrever o conteúdo de uma dada imagem, a partir das informações de formas nela existentes. Para tanto, esse método extrai inicialmente as bordas dos objetos existentes na imagem por meio do detector *Canny Edges*. Por meio de um histograma, são calculados os pontos da borda que estão no interior ou no limite da forma, determinando, assim, um vetor de pontos. Além da descrição, o trabalho propõe avaliar a similaridade entre duas formas por meio da correspondência entre pontos. Essa técnica permite reduzir o número de pontos, não sendo necessário determinar os chamados *key points* comumente utilizados em outras técnicas (BELONGIE; MALIK; PUZICHA, 2001). A subseção a seguir descreve de forma mais detalhada o método *Shape Context*.

2.3.1.1 *Shape Context*

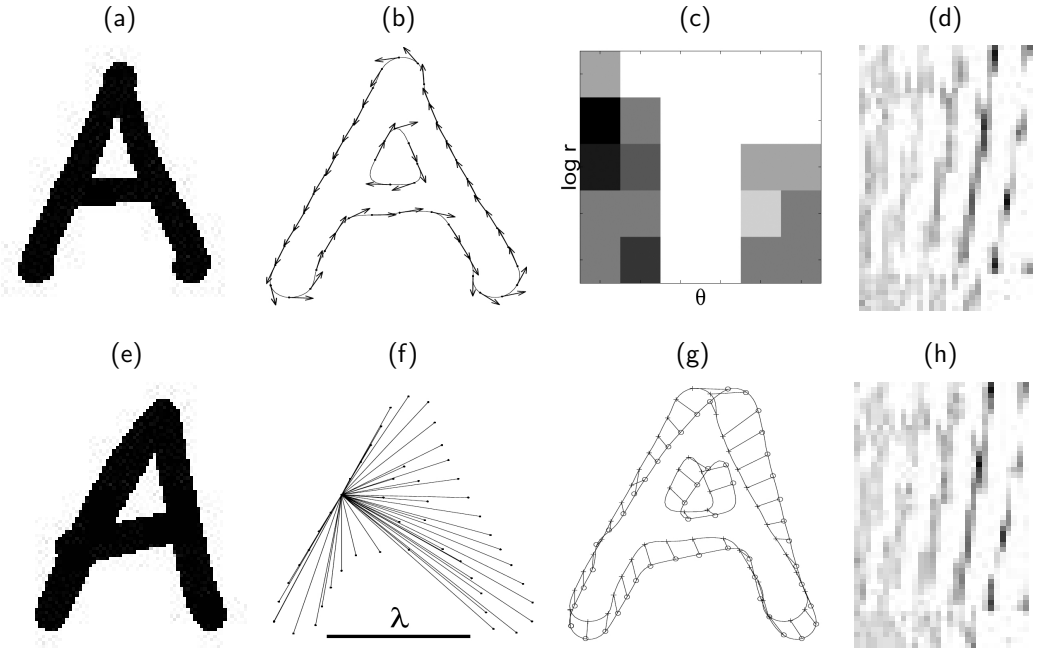
O *Shape Context* é um descritor que permite delimitar a forma, ou contorno, presente em uma imagem, com base em uma análise realizada por meio das N amostras dos pontos presentes na borda da imagem. Esses pontos podem pertencer a contornos internos ou externos de uma imagem.

É necessário que as amostras realizadas ao longo do contorno sejam uniformes em relação à sua distribuição no plano da imagem. A proposta deste tipo de descritor é delimitar a forma presente

na imagem, conforme é possível observar na Figura 17a e na Figura 17b, que corresponde ao contorno da figura anterior.

Para ilustrar o processo do *Shape Context*, e o processo de correspondência entre as bordas, tem-se que as Figuras (17a, 17e) são as imagens originais. A Figura 17b corresponde ao contorno das bordas com as tangentes delimitadas com 50 pontos de amostragem. Para cada ponto há um conjunto de vetores, que se originam a partir dele, conforme exibido pela Figura 17f, sendo que estes vetores representam a relevância do ponto. A descrição dos vetores é realizada por meio de um histograma, que permitirá distinguir de forma mais relevante a diferença entre os *pixels* nas proximidades por meio do sistema de log-polar das coordenadas. O histograma é calculado para um conjunto de vetores, como na Figura 17b, sendo o resultado mostrado pela Figura 17c, em que os *pixels* escuros denotam valores maiores. O exemplo utilizado pelos autores contêm 6 faixas (*bins*) igualmente espaçadas e com valor de 5 faixas (*bins*) para o log do raio. Com o objetivo de localizar as correspondências entre os pontos das imagens iniciais Figuras (17a e 17a), adotou-se o método húngaro, sendo considerado como pesos a dissimilaridade do histograma e a dissimilaridade do ângulo da tangente. As Figuras 17d e 17h são os resultados do processamento, formadas pelo achatamento e pelo concatenamento dos histogramas para todos os pontos das duas imagens, tendo como resultado 50 linhas, uma linha para cada ponto de amostra, e 30 colunas, uma para cada faixa (*bin*) do histograma.

Figura 17 – Ilustração do descritor e da correspondência do *Shape Context* entre imagens, as Figuras: 17a, 17e são as imagens originais; 17b contorno das bordas; 17f são os vetores de um ponto; 17c é o histograma dos vetores da Figura 17f; 17g é as correspondências; 17d e 17h são os resultados dos processamentos.



Fonte: Belongie, Malik e Puzicha (2000, p. 22)

A correspondência entre dois descritores tem que atender a dois critérios: (1) os pontos correspondentes devem ser similares; (2) as correspondências devem ser exclusivas. Logo, para calcular o custo da correspondência tem-se dois fatores: um *Shape Context* e outro local. O fator do *Shape Context* é C_S que é dado por x^2, e que corresponde à distância entre dois histogramas, representados por $k - bins$, para $g(k)$ e $h(k)$, dado pela Equação 1, com valores variando de 0 a 1:

$$C_S = \frac{1}{2} \sum_{k=1}^{K} \frac{[g(k) - h(k)]^2}{g(k) + h(k)}. \tag{1}$$

Em relação ao custo local C_A, que é uma medida do ângulo tangente da não similaridade, os valores variam entre 0 e 1, e é definido pela Equação 2:

$$C_A = \frac{1}{2} \left\| \begin{pmatrix} \cos(\theta_1) \\ \sin(\theta_1) \end{pmatrix} - \begin{pmatrix} \cos(\theta_2) \\ \sin(\theta_2) \end{pmatrix} \right\|. \tag{2}$$

Combinando os custos de correspondência de C_S e C_A por meio de uma soma ponderada: $C = (1 - \beta)C_S + \beta C_A$. Logo, o conjunto de custo é $C_{i,j}$ entre todos os pares de pontos. Os autores propuseram resolver com minimização do custo total dos pontos correspondentes, sujeito à restrição de que a combinação seja um-para-um, tendo como custo computacional resultante $O(N^3)$. Para ilustrar esse processo é apresentado pela Figura 17g.

2.4 Processamento de Linguagem Natural

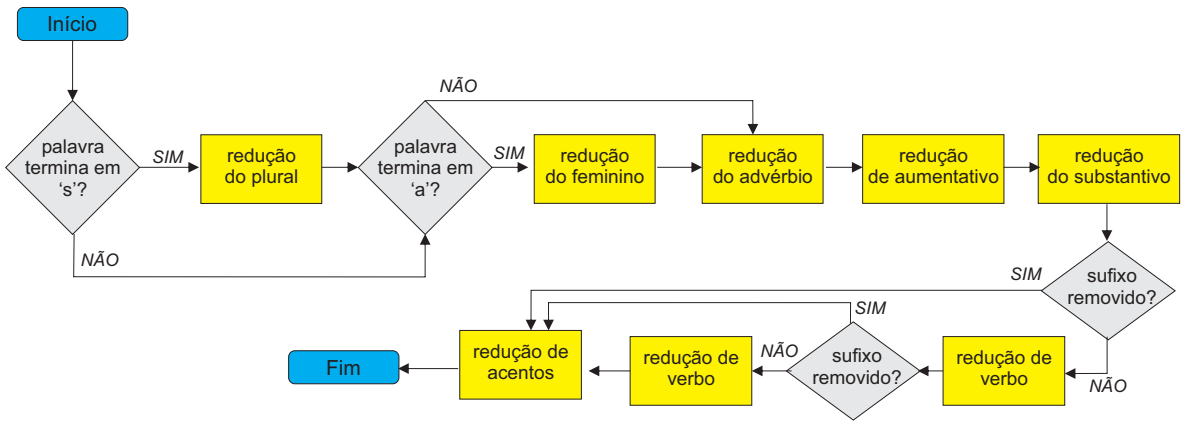
Além da visão computacional para auxiliar no processo de formação de sinais, a abordagem proposta neste trabalho baseia-se na área de Processamento da Linguagem Natural (PLN), que consiste no desenvolvimento e na aplicação de modelos computacionais para processamento de documentos que estão em alguma língua natural, como, por exemplo, um arquivo de texto qualquer. Pode-se dizer que a PLN atua em três subáreas: fonologia; morfologia e sintaxe; e semântica (ROCHE; SCHABES, 1997; MANNING; SCHÜTZE, 1999).

Em específico, pretende-se utilizar nesta tese a morfologia, que busca obter as unidades primitivas, ou raiz, das palavras. Nesse sentido, a próxima subseção irá abordar este conteúdo.

2.4.1 Extração de Radicais

A busca pelo menor elemento de um termo, ou palavra, se dá por meio da extração do radical, que permite determinar a morfologia da palavra da língua oral, tendo a língua portuguesa como referência.

Figura 18 – Fluxograma do algoritmo de Orengo e Huyck (2001).



Fonte: Adaptado de Dias e Malheiros (2005, p. 9)

O aspecto morfológico associado a um determinado termo é capturado por um processo usualmente denominado de *Stemming*, ou seja, extração de radical em português. A determinação do radical tem o objetivo de apresentar a raiz do termo, como exemplo, a palavra "terra" tem como radical "terr", as palavras "rato" e "rata" têm o mesmo radical "rat". Existe uma vasta quantidade de técnicas que permitem a extração dos radicais de palavras (ORENGO; HUYCK, 2001; DIAS; MALHEIROS, 2005; GARDIES, 2008).

Orengo e Huyck (2001) propõem um algoritmo que influenciou vários outros, sendo considerado um bom método para a determinação de radical (DIAS; MALHEIROS, 2005). O algoritmo proposto por Orengo e Huyck (2001) é representado pelo fluxograma presente na Figura 18. A execução deste algoritmo é realizada por uma sequência de passos: 1) Caso a palavra esteja no plural, faz-se a redução do plural; 2) Da mesma foram, faz-se a redução do feminino, caso a palavra seja feminina; 3) Redução do advérbio da palavra; 4) Caso a palavra contenha aumentativo ou diminutivo, faz-se também a redução destes; 5) Redução das formas nominais; 6) Remoção das terminações verbais; 7) Remoção da vogal temática e, por fim; 8) Remoção dos acentos.

Espera-se que todos os conceitos e métodos, descritos neste capítulo contribuam para a compreensão da abordagem computacional proposta neste trabalho para criação de sinais. A seguir o próximo capítulo descreverá os principais trabalhos relacionados com a presente tese.

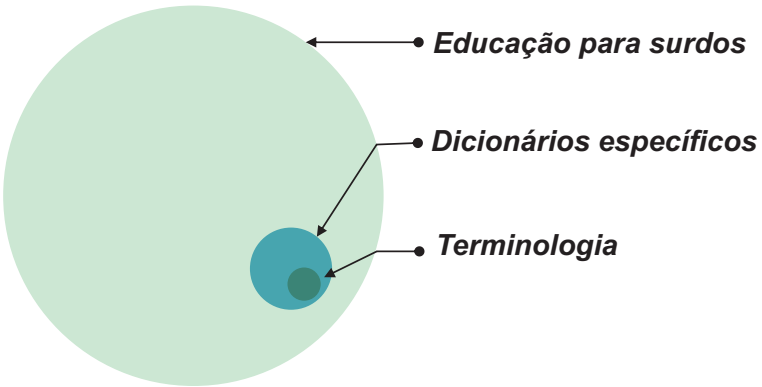
Capítulo 3

Trabalhos Relacionados

Neste capítulo são descritos alguns dos principais trabalhos da literatura relacionados ao problema de pesquisa abordado por esta tese e que contribuíram para o avanço do desenvolvimento de novos sinais técnicos para língua de sinais, em específico para as áreas de Ciência, Tecnologia, Engenharias e Matemática (CTEM).

Os trabalhos aqui apresentados são o estado da arte relacionado ao problema desta pesquisa, partindo de uma análise geral para uma análise específica do problema. Na Figura 19 é apresentada uma visão sistêmica das áreas abordadas neste capítulo, e que serão discutidas nas seções a seguir partindo de uma visão mais genérica até o cerne deste trabalho, e que, em certa medida, contribui para o avanço desta tese. Ressalta-se, que a ilustração aqui apresentada não está relacionando a importância das áreas com as dimensões apresentadas pela figura, e sim a especificidade das áreas e, consequentemente, o tema central desta tese. De certo, há muito a ser pesquisado nas áreas aqui apresentadas, como é evidenciado pela carências de propostas que desenvolvam a língua de sinais por meio da computação.

Figura 19 – Visão geral sobre as áreas a serem exploradas.



Fonte: Próprio autor.

3.1 Educação inclusiva

Nos últimos anos, várias iniciativas de pesquisa na linguística e na educação vêm buscando aprimorar o ensino do aluno surdo, bem como as línguas de sinais. Trabalhos como Roberson (2001), Stumpf (2005), Barth, Schneider e Silva (2009), Wang (2011) e Lang et al. (2007) tiveram algo em comum: a preocupação com o desenvolvimento acadêmico do indivíduo surdo.

A pesquisa apresentada por Roberson (2001) teve como foco investigar o uso da tecnologia aplicada no ensino de surdo nos Estados Unidos. Para isso, o autor entrevistou dois grupos distintos, o primeiro formado por membros da faculdade de educação para surdos, e o segundo grupo por administradores de programas de educação para surdos. Observou-se que havia um efetivo uso da tecnologia para a educação dos surdos, porém destaca-se que, por maior que seja o uso da tecnologia, os entrevistados relataram que há necessidade de uma integração maior das tecnologias nos cursos voltados para os surdos. É interessante observar que o autor, apesar de não fazer uma análise linguística, faz um direcionamento da importância do léxico da língua de sinal. Aponta também para uma necessidade de aprimoramento de estratégias em relação ao uso de computadores e de tecnologia no ensino dos surdos.

No trabalho de Stumpf (2005), um dos aspectos estudados pela autora referia-se ao letramento de crianças surdas em uma dada língua de sinais por meio do sistema de escrita *Sign Writing*. Na pesquisa, foram analisados vários aspectos relacionados à teoria da educação quanto ao processo educacional. Além de ter como base teórica o processo de alfabetização, destaca-se que esse trabalho foi realizado em dois países diferentes, Brasil e França, e, consequentemente em duas línguas de sinais distintas. Constatou-se por esse trabalho que a escrita de sinais incorporada no processo educacional das crianças surdas consolida a evolução do aprendizado.

Barth, Schneider e Silva (2009) apresentam um estudo sobre os processos e as estratégias que as crianças surdas utilizam para a construção da escrita e da leitura na língua de sinais. Os autores fazem um levantamento do marco teórico quanto ao processo de alfabetização, destacando dois aspectos nesse sentido: o primeiro está relacionado à construção do conhecimento pelo próprio indivíduo; e o segundo considera o pensamento que a criança surda realiza, no processo de alfabetização, nas atividades discursivas quanto à associação do conceito ao sinal. Esses aspectos evidenciam a importância que tem para o indivíduo surdo contar com um mecanismo de escrita no seu processo de alfabetização, o que foi constatado pela pesquisa dos autores ao introduzir um sistema computacional que escrevesse os sinais por meio da anotação do *Sign Writing*. Nessa pesquisa, concluem os autores que há uma necessidade de modificar as práticas pedagógicas da alfabetização dos surdos.

Em Wang (2011), apresenta-se uma visão geral da reforma do ensino de ciências, baseada em pesquisas realizadas nos Estados Unidos e nos seus efeitos sobre o desempenho dos alunos surdos. Depois de discutir a dificuldade de trabalhar textos científicos com indivíduos com

perda auditiva, o autor apresenta um quadro que integra a instrução baseada na pesquisa e a construção da alfabetização e seu desempenho. O autor recomenda que, para um bom desempenho no ensino de ciências, devem ser usados novos sinais com cautela, pois os alunos não podem tirar o máximo proveito sem o conhecimento necessário da língua de sinais, o que reforça a importância da abordagem proposta nesta tese, que tem a intenção de proporcionar aos alunos surdos, com os sinais apropriados, uma melhor compreensão dos conceitos importantes e complexos em CTEM.

A pesquisa de Lang et al. (2007) apresenta a experiência dos professores com o desenvolvimento de sinais e os conceitos em relação ao conteúdo, que são aspectos essenciais para a identificação de sinais adequados que garantam o aprendizado. Os resultados desse estudo indicam a necessidade de uma abordagem sistemática para avaliar a seleção do sinal e o seu impacto no processo de aprendizagem dos estudantes surdos. Relata-se que cerca de 60% dos termos considerados importantes em uma avaliação do currículo para ciências não têm representações de sinais. Nesse cenário, os autores recomendam que, à medida que esses sinais são criados, é necessário um mecanismo de avaliação crítica, o que justifica uma nova pesquisa tratando especificamente deste ponto. Eles demonstram como o processo de criação de novos sinais é complexo e, justamente por isso, devem ser realizados esforços especiais para procurar sinais técnicos de qualidade, que podem já estar em uso por usuários experientes, mas ainda não registrados em banco de dados de léxicos existentes. Neste sentido, para a educação dos surdos, faz-se necessário um conjunto de léxicos que possam transmitir os conceitos fundamentais para o aprendizado em várias áreas.

3.2 Dicionários para língua de sinais

Na área de glossários técnicos ou, mais especificamente, dicionários terminológicos para as línguas de sinais, não existem normas claras sobre o desenvolvimento de sinais para língua de sinais. O que se tem em grande maioria é a produção de pequenos glossários. O desenvolvimento de glossários, em geral, está relacionado à criação proposta entre o intérprete e o aluno surdo, e, neste caso, sofre influência de vários fatores, por exemplo, o apelo cultural/regional. Isso é, de certa forma preocupante, uma vez que a linguagem e o vocabulário são usados em várias áreas, para vários desafios em diferentes situações de leitura comuns, não só para os indivíduos com perda auditiva, mas para todos os alunos (LAMBERG; LAMB, 1980).

Uma revisão sobre os trabalhos encontrados que abordam a temática da criação de glossários para a língua de sinais e suas metodologias adotadas é realizada nesta seção com o intuito de subsidiar e de justificar o desenvolvimento desta tese e relatar as diferenças entre as abordagens apresentadas.

3.2.1 Dicionários de Propósito Geral

No que diz respeito ao desenvolvimento de dicionários de propósito geral para língua de sinais, é importante mencionar os esforços relatados por Johnston (2003), Schermer (2003), Hanke, Konrad e Schwarz (2001), Hanke e Storz (2008) e Cormier et al. (2012), bem como alguns desafios linguísticos e tecnológicos que são comuns nesse desenvolvimento.

Em Johnston (2003), apresenta-se uma proposta de banco de dados interativo, denominada "SignBank" para línguas de sinais. O SignBank permite aos usuários acessar um banco de dados de sinais do *Australian Sign Language* (Auslan), visualizar videoclipes de cada sinal individual e navegar através do banco de dados para recuperação de informações sobre as origens do sinal, seu significado, uso e gramática. Da mesma forma que os sinais do dicionário da Austrália, o banco de dados de SignBank consiste em registros de sinais sequenciados de acordo com a ordem quirêmica. O SignBank tem a capacidade de gerar automaticamente um registro de qualquer sinal recém-atestado, relatado, ou inventado e atribuir-lhe um conceito.

No trabalho de Schermer (2003) é descrito o processo de padronização do léxico e da gramática da Língua de Sinais dos Países Baixos, *Sign Language of the Netherlands* (SLN) e o desenvolvimento de novos sinais a serem utilizados nas escolas por alunos surdos. Duas iniciativas principais se destacaram neste sentido: *Kommunikatieve Vaardigheden* (KOMVA), competência comunicativa, projeto de 1982 - 1990, e do *Standardization of Basic Lexicon* (STABOL), Padronização de Lexicon Básico, projeto de 1999 - 2002. Para conceitos mais específicos, como o vocabulário para matérias ensinadas na escola (por exemplo, geografia e matemática), o autor relata que menor variação regional foi observada e novos sinais muitas vezes tiveram que ser desenvolvidos.

Hanke, Konrad e Schwarz (2001) apresentam uma ferramenta denominada "GlossLexer", criada para apoiar a produção de dicionários de língua de sinais, tanto em versão impressa como em versões multimídias, derivadas de transcrições de declarações de linguagem de sinal. Esse instrumento foi baseado em um banco de dados lexicais, que foi continuamente aumentado com as transcrições. Infelizmente, a ferramenta GlossLexer não conseguiu descrever declarações com sinais complexos, devido à sua abordagem que reduziu esses sinais a sequências de entidades lexicais muito simples. Para superar essa limitação, Hanke e Storz (2008) propuseram a ferramenta alternativa chamada Ilex, que combina um léxico e uma base de dados de transcrição para a linguagem de sinais em toda a sua complexidade. Em Ilex transcrições consistem em fichas, ou seja, exemplos de ocorrências de sinais que fazem referência aos respectivos tipos. Como argumentado pelos autores, os dados lexicais tem uma relevância imediata para o processo de lematização. A ferramenta Ilex suporta as tecnologias da Web 2.0 para o trabalho colaborativo. As transcrições das fichas são recursos que trazem, no entanto, um desafio de estabelecer convenções de transcrição entre diferentes usuários.

Cormier et al. (2012) apresentam o desenvolvimento de um dicionário *on-line*, denominado BSL SignBank, originado a partir de uma base de dados lexical em BSL, previamente construído para estudar algumas questões em relação às frequências lexicais. Nesse trabalho, os autores descrevem os principais critérios utilizados na tomada de decisões relacionadas com, por exemplo, lematização e forma de citação com base em conjuntos de variantes fonológicas.

No trabalho de Kanis et al. (2011), os autores apresentam um dicionário de linguagem de sinais desenvolvido por uma equipe de pesquisa envolvendo a Universidade de West Bohemia, a Universidade de Masaryk e a Universidade Palacky. O dicionário foi projetado para uma aplicação *on-line*, com uma estrutura hierárquica para garantir a consistência lexicográfica do conteúdo do dicionário. A proposta dessa aplicação é de representar as línguas faladas e as várias representações formais para línguas de sinais, tais como: vídeos, imagens, HamNoSys, SignWriting e um avatar 3D interativo.

Finalmente, os autores em Vogler e Goldenstein (2008) descrevem alguns dos problemas atuais no processamento computacional de línguas de sinais, incluindo o elevado nível de ações simultâneas, a interseção entre sinais e gestos e a complexidade de modelagem de processos gramaticais. Os autores argumentam que, ao contrário de línguas faladas, não há consenso sobre o que exatamente é a estrutura básica de um sinal, como evidenciado por diferentes modelos fonológicos na literatura linguística (BRENTARI, 1998). Apontam para o objetivo de dividir os sinais em partes menores de uma unidade léxica, tanto pelo meio computacional como pela análise linguística, sendo este um caminho promissor de pesquisa.

3.2.2 Dicionários específicos

Esforços recentes têm abordado os problemas do desenvolvimento de glossários terminológicos para a língua de sinais, principalmente, para a CTEM. Alguns projetos, por meio de uma análise cuidadosa, realizada por comitês de especialistas ou por trabalho colaborativo, envolvendo diferentes pessoas da comunidade surda, resultam na criação de dicionários de língua de sinais que são disponibilizados *on-line* ou, por exemplo, em outras mídias, tais como DVD (BIGHAM et al., 2008; CAVENDER; LADNER, 2010; HOLGATE, 2015).

Em Bigham et al. (2008) e Cavender et al. (2010), por exemplo, os autores descrevem a experiência do Fórum ASL-STEM (LADNER; CAVENDER, 2009), que utiliza recursos de vídeo e interação entre os usuários, oferecendo ainda um local para colaboração para a linguagem e discussão sobre *American Sign Language* (ASL), em relação à terminologia STEM (BIGHAM et al., 2008). Segundo os autores, o Fórum ASL-STEM obteve uma rápida aceitação da comunidade surda, por desenvolver sua plataforma pensando nas questões dos vídeos transmitidos, acessados ou disponibilizados via smartphone, vlogs (blogs de vídeo) orientados aos surdos, e compartilhamento nas redes sociais por meio de vídeos. As classificações da comunidade são usadas para enfatizar os sinais mais bem cotados para cada conceito, utilizando

uma abordagem de discriminação colaborativa (KONSTAN et al., 1997).

Em Holgate (2015), são discutidas algumas iniciativas na Língua Britânica de Sinais, em inglês *British Sign Language* (BSL), como o Projeto de Glossário do Centro Sensorial Escocês (CSE). Esse trabalho visa desenvolver listas de termos específicos de um assunto em BSL, promovendo as definições em filmes e um laboratório para adicionar novos conhecimentos nas definições dos termos já existentes. O público-alvo é constituído basicamente por jovens surdos que estão na escola, que utilizam BSL, e querem aprender de forma independente, usando a Internet. Novos sinais são frequentemente criados por uma equipe de especialistas (cientistas surdos e matemáticos ou professores de ciências), utilizando informações sobre o conceito e seguindo os princípios do léxico produtivo da BSL.

Ao se analisar na ASL trabalhos que justificam o desenvolvimento de novas terminologias para a CTEM, tem-se o trabalho de Kelly e Gaustad (2006) e posteriormente são desenvolvidos dicionários nos trabalhos de Vesel e Robillard (2013a) e Vesel e Robillard (2013b).

Em Kelly e Gaustad (2006), os autores examinam as relações específicas entre o desempenho dos estudantes universitários surdos em matemática e suas habilidades em leitura, linguagem e morfologia em Inglês. Os autores argumentam que, para uma boa interpretação do conteúdo, é necessário conhecimento da língua inglesa, seja oral ou impressa, além da língua de sinais. Sendo assim, é necessário um maior desenvolvimento do vocabulário na modalidade da língua de sinais, pois há uma carência neste aspecto. Nesse caso, novos sinais permitiriam diferenciar melhor os conceitos sinônimos, assim como codificar conceitos complexos sem recorrer à datilologia. Segundo os autores, a criação de um sinal deve considerar o componente de natureza morfêmico das palavras, no caso da língua inglesa, utilizadas no texto para representar a língua de ensino correspondente em sala de aula. Esta observação corrobora para o uso, no âmbito da abordagem proposta, de um método de processamento típico de linguagem natural (decorrente do algoritmo) para determinar a conexão dos termos inexistentes no léxico da língua de sinais considerada, com o intuito de criar automaticamente um sinal com representação correspondente.

Vesel e Robillard (2013a) descrevem o uso do Desenho Universal para Aprendizado (DUA), uma abordagem para apoiar a criação de dicionários em CTEM para o ASL. Os dicionários criados contêm aproximadamente 750 termos e definições para conteúdos específicos, a maioria dos quais incluindo uma ilustração ou exemplo, e utilizam personagens virtuais para a apresentação dos sinais. Os dicionários propostos pelos autores têm servido para padronizar os sinais utilizados pelas escolas e pelos intérpretes que não têm um boa formação sobre os vocabulários em CTEM. Da mesma forma, em Vesel e Robillard (2013b), é relatada uma experiência com a construção de um dicionário de matemática, nomeado Dicionário de Sinais para Matemática (DSM) a fim de ajudar os alunos surdos na escola com o vocabulário dessa área. Os resultados desse estudo sugerem que, quando utilizado em ambientes reais, como sala de aula, o DSM

pode contribuir de forma significativa para um processo de aprendizagem efetiva dos alunos surdos.

3.3 Formação das palavras

No âmbito do processo de formação das palavras, a quantidade de trabalhos para a língua de sinais é bem menor do que para a língua escrita. Entre eles podemos destacar especificamente: Bernardino (1999), Quadros e Karnopp (2004), Stumpf (2005), Felipe (2006), Barth, Schneider e Silva (2009), Nascimento (2009), Takahira (2012) e Lima (2014).

3.3.1 Morfologia tradicional

O processo da análise das palavras nas línguas faladas é denominado morfologia. A análise morfológica realizada na língua portuguesa, por exemplo, permite extrair diversas informações que possibilitam classificar e, principalmente, informar sua origem. Em relação à língua de sinais, os trabalhos abaixo relacionados realizam uma reflexão sobre esse tema, uma vez que não existem regras bem definidas relativas ao processo de criação dos sinais (FELIPE, 2006; TAKAHIRA, 2012).

Bernardino (1999) discute a morfologia do sinal em Libras, que sofre grande influência do fato de ser uma língua multidimensional, em que os parâmetros podem variar em relação ao aspecto, à inclusão de quesitos gramaticais e lexicais. No estudo, a autora observou que os sujeitos analisados não utilizam propriamente a Libras. A língua utilizada pelos falantes da comunidade surda adulta é uma língua que possui características próprias, definidas pelos usuários. Os sujeitos analisados relatam que há uma criação/construção da língua, mas ressaltam que, ao entrarem em contato com os fluentes em Libras, podem tornar a reconstruir os sinais já criados. A conclusão da autora é a de que os sujeitos analisados utilizam uma variação da Libras, denominada "crioula", para se comunicarem e aponta como possível causa para isso a forma de acesso à língua.

Sobre o aspecto da morfologia tradicional, Quadros e Karnopp (2004, p.94) fazem uma reflexão sob a ótica da língua falada abordando duas áreas: a derivacional e a flexional. A autora apresenta as influências da análise morfológica dos sinais em relação à língua falada, estabelecendo um certo grau de aproximação.

O trabalho apresentado por Stumpf (2005) mostra que a Libras possui todas as características de língua, como: fonologia (configuração de mãos, alfabeto, uso de uma ou duas mãos, movimento), morfologia (plural, intensidade, modo, tempo, forma, tamanho, classificadores, incorporação e negação), sintaxe (exploração do uso do espaço, organização de objetos e referentes presentes e não-presentes, concordância nos verbos, marcação dos verbos sem concordância, estruturas complexas, topicalização), semântica (emprego das relações do significado no nível lexical:

antonímia, sinonímia, homonímia, polissemia) e pragmática (exploração de implicaturas, figuras de linguagem, formas de polidez na linguagem), concluindo, assim, que é uma língua definida como qualquer outra.

No artigo de Felipe (2006), são analisados os processos de formação de palavras na Libras. Para a autora, são considerados como morfemas os parâmetros como: configuração de mão, direcionalidade, ponto de articulação, movimento, localização do toque, expressões faciais, constituindo-se, assim, um conjunto de sistemas complexos que, em seus processos, contribuem para a formação de novos sinais. A Libras é considerada uma língua flexional, mas comporta também características de língua aglutinante, as quais são perceptíveis pelos processos de composição e de incorporação; apresenta, ainda, característica de língua classificadora, a qual ocorre como acréscimo a um radical nominal ou como uma modificação interna da raiz verbal do sinal. O trabalho ressalta que não tem como objetivo definir os processos de criação de sinais.

Nos resultados apresentados por Barth, Schneider e Silva (2009) foi possível observar que uma das crianças surdas que estava participando dos experimentos passou a utilizar o teclado virtual para a escrita de sinais em Libras no processo de construção de um novo vocabulário. Foi observado que o indivíduo passou a adotar pequenos sinais que representavam objetos em frases curtas, com uma compreensão maior da conversação entre os pares. Os autores observaram ainda que as crianças criam seus próprios mecanismos durante o processo de evolução da escrita da língua de sinais. Por fim, ressaltam que é de grande importância o desenvolvimento de metodologias de construção da escrita e da leitura, na língua de sinais, mais adequadas ao desenvolvimento linguístico para crianças surdas.

Nascimento (2009), em seu estudo sobre a representação léxico da Libras, descreve que o lexical ocorre de forma ordenada com a combinação de características morfológicas da Libras. Diferente dos outros trabalhos, a autora realiza uma descrição mais aprofundada sobre as características morfológicas da Libras fundamentando-se em duas áreas: lexicologia, que aborda a análise teórica da categorização, os processos de constituição e da construção do léxico, da semântica dos protótipos, dos classificadores; e lexicografia, que realiza análise de dicionários existentes, pressupostos teóricos, e da representação iconográfica. Como conclusão desse trabalho, a autora apresenta três propostas: uma lexicográfica, uma terminográfica e uma lexicográfica teórica. Porém, tais propostas estão limitadas a uma determinada realidade e não concluem os modelos propostos, apenas apresentam um direcionamento para o desenvolvimento de trabalhos relacionados à morfologia em Libras.

É observado por Takahira (2012) que existe uma carência muito grande sobre os sinais compostos e as questões morfológicas relacionadas aos sinais em Libras, campos em que a maioria dos estudos não consegue descrever os fenômenos da língua. É possível descrever, em relação aos sinais compostos em Libras, que existem diferentes tipos, como os de justaposição

e os de aglutinação, que em geral são relacionados às línguas orais. Este trabalho evidencia que a literatura atualmente disponível, principalmente em relação à morfologia na Libras, não é capaz de explicar esses fenômenos.

Com base no estudo dos trabalhos apresentados, os quais representam, em certa medida, o estado da arte no processo de formação de sinais no âmbito da língua brasileira de sinais, em especial o processo de associação cognitiva, é possível observar uma forte influência da iconicidade sobre a geração de novos sinais em relação à língua oral; no entanto esta não é considerada a única influência, o que será explanado nas próximas subseções. Observa-se ainda que os trabalhos nessa área de pesquisa debruçam-se sobre estudos e análises em *corpus* significativamente limitado, visto que todo o processo de formação de sinais não apresenta um conjunto de métodos definidos.

3.3.2 Iconicidade

A iconicidade é uma propriedade cognitiva usada principalmente na linguística (NUNES, 2009; SIMÕES, 2009; COSTA, 2012; TEMOTEO, 2012; GUIMARÃES, 2013). Ela permite associar um ícone ou um signo icônico a um conceito relativo a algo real e, por isso, tem grande influência na formação de sinais em línguas visoespaciais. Nesta subseção contextualiza-se como os sinais sofrem a influência dessa propriedade a qual, em certa medida, também influencia a criação de novos sinais.

A pesquisa de Nunes (2009) realiza um levantamento simplificado sobre a análise da língua de sinais segundo a Linguística Cognitiva. Em geral, conforme destaca o autor, existem restrições do uso do alfabeto manual pelos surdos tendo como preferência o uso dos sinais. De fato, os sinais apresentam uma relação com a linguagem corporificada. Essa linguagem corporificada são sinais que, em geral, são realizados próximos a partes do corpo com funções específicas, como exemplifica a autora no caso do verbo chorar, para o qual o sinal é apresentado próximo aos olhos. Há uma discussão de diversos autores em relação ao posicionamento de linguistas sobre a iconicidade em língua de sinais conforme destaca a autora.

No trabalho de pesquisa de Costa (2012), é explorada a questão da produtividade da língua de sinais. Em geral, o princípio da dupla articulação da linguagem na criação do sinal tem como base três parâmetros: configuração de mão, movimento e locação, os quais, quando isolados, não têm significado, mas, quando combinados, proporcionam o desenvolvimento de novas unidades com significado e que, conforme a autora, descrevem novos morfemas. A proposta do trabalho de Costa (2012) é que esses parâmetros sofrem influência da iconicidade, e os resultados sugerem que essa influência é algo realmente presente na língua de sinais.

Costa (2012) demonstra que, em línguas de sinais, criam-se novos signos pela arbitrariedade, não somente quando há falta de motivação, mas também quando ocorre a impossibilidade

de associação a um signo linguístico. Destaca-se ainda que a iconicidade está presente nas línguas de sinais, em grau mais elevado do que nas línguas faladas, levando-se em consideração o fato de que línguas de sinais são de característica visual, acarretando assim, uma maior busca por ítens icônicos. O autor destaca ainda que a análise na língua de sinais é estabelecida em dois eixos. Um associado aos estudos das línguas de sinais em relação aos parâmetros, como, a configuração de mão em relação à sua formação, e a influência da iconicidade ou da arbitrariedade. E, um segundo eixo que tem como base as línguas naturais, tendo como foco a arbitrariedade e a iconicidade nas línguas de sinais.

Guimarães (2013) apresenta uma proposta de arquitetura computacional pedagógica para letramento bilíngue de surdos. Essa arquitetura baseia-se nas características das línguas de sinais com grande influência da iconicidade, incorporando o uso de teorias e ferramentas capazes de alterar as estruturas cognitivas, tais como Mapas Conceituais, *Sense-Making*, e Senso Comum. Os principais usuários da arquitetura são os educadores e outros membros da comunidade surda, sendo voltada para o propósito de promover o letramento bilíngue e as interações intelectuais, no que se refere a interações sociais mediadas por artefatos computacionais que instigam o uso de processos intelectuais para o desenvolvimento do conhecimento e de amadurecimento em geral.

Os autores Perniss e Vigliocco (2014) apresentam uma visão teórica quanto à iconicidade na BSL. Eles partem da compreensão do processamento, do desenvolvimento e da evolução da linguagem e questionam a visão de muitos autores, que consideram sua natureza arbitrária como uma característica importante do sistema linguístico humano. Logo, eles demonstram que a iconicidade na comunicação cara a cara é um meio importante, e que essa fornece um papel ímpar para a compreensão e para a evolução da linguagem. Ressalta-se que no processamento de linguagem, a iconicidade pode fornecer um mecanismo para explicar o modo como ela se concretiza, sendo fundamental para uma comunicação significativa.

Na pesquisa de Perlman, Dale e Lupyan (2015), os autores propõem uma avaliação da comunicação por meio de uma linguagem rudimentar para verificar a relação com a iconicidade. Os autores examinaram o processo de desenvolvimento da língua, na qual pares de participantes (surdos e ouvintes) desenvolveram novos sinais para 18 termos, por meio de um jogo de comunicação de charadas interativo. Foi observado que os surdos convergiram rapidamente para sinais estáveis e os ouvintes ingênuos podiam inferir corretamente seus significados. Logo, pode-se concluir que a capacidade das pessoas de adivinhar os significados ocorreu devido à proximidade dos sinais com um "modelo de significado" icônico. Sendo assim, os resultados sugerem que a significância desses sinais se originou da iconicidade.

Para Stumpf (2005), as crianças de um ano e meio a três anos estão no auge do aprendizado, ou seja, do poder cognitivo. O Cognitivismo possui a característica de assimilação definido por meio de representações de ícones para os conceitos.

Um dos objetivos desta tese é adotar um processo de criação de sinais, que em certa medida, está relacionado à iconicidade. A iconicidade muitas vezes é tratada como um aspecto cognitivo do indivíduo que se vale dela para realizar uma ligação entre um conceito e uma imagem. Sendo assim, cada palavra criada sobre o aspecto da iconicidade tem uma ligação cognitiva a representação gráfica. Esta associação entre conceito e imagem permitirá à abordagem proposta neste trabalho para criação de sinais adotar, em certa medida, este aspectivo cognitivo, de relacionar novos conceitos a imagens. Contudo, sabe-se ainda pouco sobre como as pessoas criam os sinais, e muitos sugerem que vários sinais não fundamentam-se na iconicidade além das relações triviais de onomatopeia (PERLMAN; DALE; LUPYAN, 2015). Dessa forma, se faz necessário também avaliar outras características, como a arbitrariedade.

3.3.3 Arbitrariedade

Um dos processos de formação das palavras em línguas de sinais é o da arbitrariedade, descrito por vários autores como um processo arbitrário sem um único traço característico de formação (BERNARDINO, 1999; PIVETTA; SAITO; FLOR, 2013; FRYDRYCH, 2012; GOMES; TEIXEIRA, 2014).

No estudo de Silva (2009), são utilizadas outras línguas para a análise linguística da língua de sinais. Considera-se que não há prejuízo nessa metodologia, principalmente em relação à análise quanto ao seu *status* de língua natural. Essa preocupação quanto ao *status* de língua natural é comum na área, tanto que as análises buscam por traços que contenham propriedades linguísticas como: arbitrariedade; flexibilidade e versatilidade; descontinuidade; criatividade/produtividade; dupla articulação; padrão; e dependência estrutural (SILVA, 2009, p.17).

Para Nunes (2009), o processo de sinais por meio da iconicidade não define a sua formação, e causa empecilhos para a análise dos sinais. Destaca que a arbitrariedade é uma característica presente nas línguas naturais. As línguas são arbitrárias, pois os sinais têm uma relação arbitrária entre a forma e o significado, sendo impossível prever o significado pela forma e o contrário também é verdade, pois dado o significado não é possível prever a forma.

Frydrych (2012) realiza uma análise a respeito dos processos de formação das palavras pela arbitrariedade e pela iconicidade. A hipótese do autor é de que a influência inicial é reforçada pelo seu estatuto linguístico, tendo como fonte trabalhos clássicos na área da linguística. Para o autor, o processo arbitrário tem como base o fundamento saussuriano, que diz que:

> "não deve dar a ideia de que o significado dependa da livre escolha do que fala (ver-se-á, mais adiante, que não está ao alcance do indivíduo trocar coisa alguma num signo, uma vez que esteja ele estabelecido num grupo linguístico); queremos dizer que o significante é *imotivado*, isto é, arbitrário em relação ao significado, com o qual não tem nenhum laço natural na realidade" (SAUSSURE, 2006, p.83 - grifos do autor).

O fundamento saussuriano apresenta os conflitos em separar a língua de sinais da língua falada, concluindo ainda que a língua é dada como produto do próprio indivíduo. Para o sinal não existe a obrigatoriedade de seu registro como na língua oral, não sendo possível, em grande parte dos casos, a sua preservação.

Em contrapartida, Frydrych (2012) diz que a fala é a ação do próprio indivíduo, podendo ser distinta pelos arranjos em que o falante apresenta um código da língua com o objetivo de demonstrar seu pensamento.

Apesar de muitos conceitos trabalhados pelo autor não estarem direcionados à língua de sinais, é possível realizar essa relação, pois há subsídios para isso, uma vez que os princípios que regem esses sistemas são comuns às línguas. O conceito de arbitrariedade é também compreendido como a ausência de causalidade, não tendo um referencial ou uma regra para a formação ou, até mesmo, a união de signos linguísticos.

Sobre a arbitrariedade, Gomes e Teixeira (2014) destacam que, na constituição do signo linguístico, a relação entre o significante e o significado é arbitrária, não tendo nenhuma relação entre a forma do significante e o seu conteúdo (significado), sendo este conceito advindo do fundamento saussuriano.

O trabalho de Tolar et al. (2008) teve como objetivo medir a compreensão das crianças sobre a relação entre o sinal e a iconicidade. Para avaliar tal relação, o trabalho abordou, por meio de uma pesquisa de entrevista com 64 participantes de idades de 2,5 a 5 anos que foram agrupados de acordo com a idade. Os participantes não tinham exposição prévia aos sinais pesquisados, devendo responder à combinação entre os sinais icônicos com imagens de referência. O resultado obtido por esta pesquisa é que, à medida que os participantes eram mais velhos, gradualmente aumentava a porcentagem em que reconheciam a relação do sinal icônico com a imagem. Isso se deve ao fato de que o desenvolvimento do indivíduo ocorre nos anos pré-escolares. Pode-se concluir por esse trabalho que a relação da iconicidade está relacionada ao meio educacional do indivíduo; logo, a relação entre o signo e o significado pode ficar comprometida, sendo arbitrária a relação existente.

Para Nascimento (2009) a arbitrariedade do signo linguístico vem sendo compreendida entre os pesquisadores como a ausência de algo que permita referenciar a forma de um sinal e o seu significado, e essa ideia é considerada como algo irrefutável por muitos linguístas. Não há, portanto, nesse caso, um procedimento metodológico para a criação de um sinal.

Finalmente, assumindo-se que os sinais não são criados apenas com base na iconicidade, mas também arbitrariamente, o que consiste na convenção do indivíduo ao propor um novo sinal, propõe-se neste trabalho o uso de um método típico de processamento de linguagem natural (ORENGO; HUYCK, 2001) para determinar o radical do termo associado ao conceito sem

sinal na língua oral de referência do surdo (por exemplo, português, no caso da Libras). Este radical é posteriormente utilizado para procurar modelos de sinais com natureza morfêmica similar (radical similar). A ideia é que essa pista alternativa possa contribuir para propor um sinal significativo, especialmente quando as características icônicas não podem ser estimadas com precisão (conceitos abstratos).

3.3.4 Demais Processos Linguísticos na Formação de Sinais

É importante ressaltar que todos os trabalhos apresentados anteriormente, discutiram a importância da análise linguística na língua de sinais, bem como os processos que influenciam a sua formação. De fato, o processo de formação de sinais ainda não é algo plenamente consolidado.

Lima (2014) desenvolveu um dicionário terminológico para a área do desenho arquitetônico para Libras, que teve como referência as diretrizes baseadas na análise da língua oral. Nesse trabalho, foram descritos todos os processos de criação dos sinais, como a iconicidade e a arbitrariedade. Essas descrições permitem analisar e avaliar a influência dos processos linguísticos envolvidos na formação de sinais. Por meio de fichas lexicográficas, contendo diversos dados sobre os sinais, foi gerado um repositório terminológico do desenho arquitetônico, contendo 74 termos que foram catalogados. A Tabela 3 é uma compilação dos dados coletados nas fichas dos termos desenvolvidos.

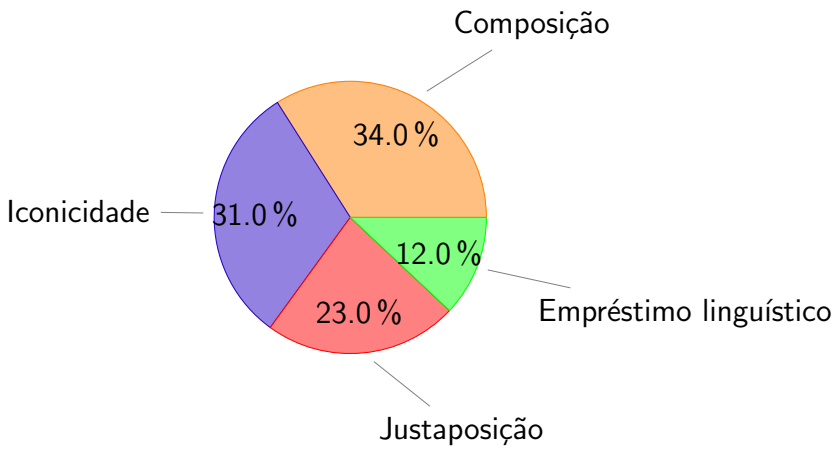
Observa-se, a partir da Tabela 3, que há predominância de mais de um processo linguístico na formação dos sinais. Especificamente, 62% das formações ocorreram sob a influência de dois ou mais processos de formação. Esse resultado sugere que a proposta apresentada por esta tese é factível, qual seja, a de adoção de dois processos para a geração de sinais: o primeiro baseado na iconicidade e/ou na arbitrariedade; e o segundo, em um processo de composição fundamentado nas características de forma extraídas de símbolos em SW.

De fato, em Lima (2014), o estudo realizado evidenciou que os processos de formação de sinais ocorrem, predominantemente, por composição e por iconicidade, conforme é possível observar na Figura 20. Esses dois processos correspondem a mais da metade das formações. A justaposição, que corresponde a 23%, ocorre na criação de termos compostos, por exemplo, "sala de estar". Já o empréstimo linguístico ocorre, com uma frequência menor, para descrever

Tabela 3 – Número de processos linguísticos envolvidos na formação dos sinais propostos no trabalho de Lima (2014).

Número de processos linguísticos	**Quantidade de Sinais**	**%**
1	28	38,00
2	29	39,00
3	17	23,00

Figura 20 – Gráfico da análise do processo de formação de sinais.



termos, como por exemplo, "engenharia de produção".

3.4 Discussão

Considerando o estado da arte relativo ao desenvolvimento de soluções para o problema da escassez de léxicos terminológicos em línguas de sinais, nota-se que os avanços obtidos foram pontuais e em alguns casos, resolveram problemas locais. A necessidade da evolução da língua de sinais é evidenciada pela Seção 3.1, na qual há ideias/teorias de vários autores que apontam para necessidade de novos termos para auxiliar na educação dos surdos, e, ainda, que a criação de conjuntos de termos específicos poderiam contribuir de forma significativa.

Ao se analisar os dicionários disponíveis para as línguas de sinais existentes, pode-se concluir que são poucos e com um número pequeno de termos se comparados com os das línguas orais. Analisando-se que o processo de construção de dicionários para língua de sinais, os trabalhos argumentam sobre a dificuldade e a falta de mecanismos que permitam os seus desenvolvimentos. São apresentados nas subseções 3.2.1 e 3.2.2 alguns trabalhos que discutem esta questão e suas soluções. A Tabela 4 consiste no resumo de alguns trabalhos, sendo analisados quanto à proposta e sua crítica.

A maioria dos trabalhos apresentados pela Tabela 4 tem a língua ASL como objeto de estudo, podendo destacar os trabalhos de Bigham et al. (2008) e Cavender e Ladner (2010), os quais desenvolvem um dicionário colaborativo com avaliação dos termos pelos usuários. A proposta é inovadora e aborda a questão da rede como uma ferramenta importante na construção de um dicionário, porém em certa medida não há controle efetivo dos termos e, considerando os aspectos que descrevem o sinal, fica a cargo do colaborador fazer as anotações. Semelhante problema ocorre no trabalho de Vesel e Robillard (2013b). Outra característica que se destaca, está relacionada à proposta do glossário desenvolvido e que, neste caso, há dois trabalhos que têm a matemática como tema, o de Kelly e Gaustad (2006) e o de Vesel e Robillard (2013b),

Tabela 4 – Resumos das principais obras que desenvolveram glossários gerais e específicos.

Trabalho	**Proposta**	**Crítica**
(KELLY; GAUSTAD, 2006)	Análise e desenvolvimento de sinais.	Não foi proposta nenhuma metodologia para auxiliar a criação.
(BIGHAM et al., 2008)	Fórum ASL - Term - Colaborativo.	Não há controle sobre os sinais.
(CAVENDER; LADNER, 2010)	Fórum ASL - Term - Colaborativo e classificação dos sinais.	As transcrições das fichas são um importante recurso, no entanto, há o desafio de estabelecer convenções de transcrição entre diferentes usuários.
(VESEL; ROBILLARD, 2013a)	Desenho Universal para Aprendizado - STEM.	Sem um controle efetivo dos sinais e dos processos.
(VESEL; ROBILLARD, 2013b)	Desenho Universal para Aprendizado - Matemática.	Sem um controle efetivo dos sinais e dos processos.
(HANKE; KONRAD; SCHWARZ, 2001)	Apoiar a produção de dicionários de língua de sinais por meio de anotações.	Entidades lexicais muito simples e sem controle das inclusões.
(JOHNSTON, 2003)	Banco de dados interativo de sinais.	Sem controle do sinais inseridos.
(SCHERMER, 2003)	Processo de padronização do léxico e da gramática.	Há a presença da variação regional e novos sinais muitas vezes tiveram que ser desenvolvidos.
(HANKE; STORZ, 2008)	Combina léxico e uma base de dados de transcrição.	As transcrições das fichas são um importante recurso, no entanto, há o desafio de estabelecer convenções de transcrição entre diferentes usuários.
(CORMIER et al., 2012)	Dicionário *on-line* focado com a lexicografia.	Termos desenvolvidos de forma manual.

Fonte: Próprio autor.

ou seja, duas fontes de informações para a mesma língua, o que pode ser um fator preocupante para a aceitação dos sinais desenvolvidos. No trabalho de Cormier et al. (2012), os autores tiveram a preocupação de tratar a língua de sinal com elementos descritivos da linguística, porém a sua construção não é rápida.

Considerando esses glossários apresentados, observa-se que há uma carência por falta de processos que definam o sinal desenvolvido, e que muitas vezes o processo colaborativo sem um controle pode prejudicar o desenvolvimento da língua, como relatado nos trabalhos de Schermer (2003), Cavender e Ladner (2010) e Vesel e Robillard (2013b). Sendo assim esta característica de padronização foi desenvolvida por este trabalho ao utilizar a computação para auxiliar no desenvolvimento de novas terminologias para a língua de sinais de forma robusta e padronizada.

A criação de neologismos é algo que a humanidade vem realizando ao longo dos tempos, e que ultimamente recebe a colaboração da computação (STEELS, 2011; WELLENS et al., 2013). A adoção de máquinas nesta tarefa é cada vez mais comum, porém é necessária toda uma análise do processo de formação da palavra. É neste sentido que foram apresentados na Seção 3.3 os aspectos a serem considerados na construção do sinal. Considerando os trabalhos expostos neste viés, a morfologia do sinal é caracterizada como a língua oral, ao considerar várias unidades da linguística no processo construtivo, destacando-se os trabalhos de Quadros e Karnopp (2004) e Felipe (2006), que fazem uma imersão neste assunto.

De fato, a morfologia permite compreender os aspectos construtivos tanto das palavras quanto dos sinais, mas a origem de sua formação é tratada por outros elementos relacionados à cognição humana, ou seja, iconicidade e arbitrariedade. Ao tratar dessa temática, a literatura não conclui qual é o processo que origina a formação dos sinais, há sempre um estudo que comprova a hipótese defendida, como o trabalho de Perlman, Dale e Lupyan (2015) que defende a iconicidade; e de Gomes e Teixeira (2014) que defende a arbitrariedade. Há uma grande discussão acerca desta temática, e esta tese não tem como foco definir qual é realmente a influência de criação dos sinais, e sim reconhecer a importância de tais hipóteses e modelá-las computacionalmente.

Dado o contexto acima descrito, a presente tese aborda o problema da ausência de terminologia em línguas de sinais para conceitos técnico-científicos, por meio de uma metodologia computacional inédita, capaz de apoiar a criação de dicionários terminológicos para diferentes áreas em CTEM. Vale ressaltar que, segundo dados do censo demográfico de 2010 do Instituto Brasileiro de Geografia e Estatística (IBGE), 9,7 milhões de brasileiros eram detentores de algum tipo de deficiência auditiva, dos quais cerca de 350 mil eram pessoas surdas (IBGE, 2010). Tais cidadãos, em sua vasta maioria, estão envolvidos em trabalhos exclusivamente operacionais, os quais estão sabidamente em processo de extinção com os novos avanços tecnológicos.

Neste cenário, esta tese orienta-se, não somente pela fluidez das mudanças econômicas e sociais, características dos tempos atuais, mas também se fundamenta na robustez de uma abordagem inovadora, proposta para desenvolvimento de linguagem. Sendo assim, este trabalho foi conduzido sob a perspectiva de uma pesquisa interdisciplinar, transitando entre as áreas da computação e linguística, para cumprir o seu propósito de qualificar o surdo para além do trabalho operacional, principalmente nas áreas de CTEM.

Capítulo 4

Abordagem Proposta

Neste capítulo, será descrita a abordagem computacional proposta, que permitirá dar suporte à criação de novos termos para dicionários terminológicos, em língua de sinais, voltados para áreas em CTEM. Conforme ilustrado pela Figura 21, o processo de geração de sinais consiste em cinco etapas:

1 Classificação de modelos de sinais por similaridade de ilustrações;

2 Classificação de modelos de sinais por similaridade de radicais de termos na língua oral de referência;

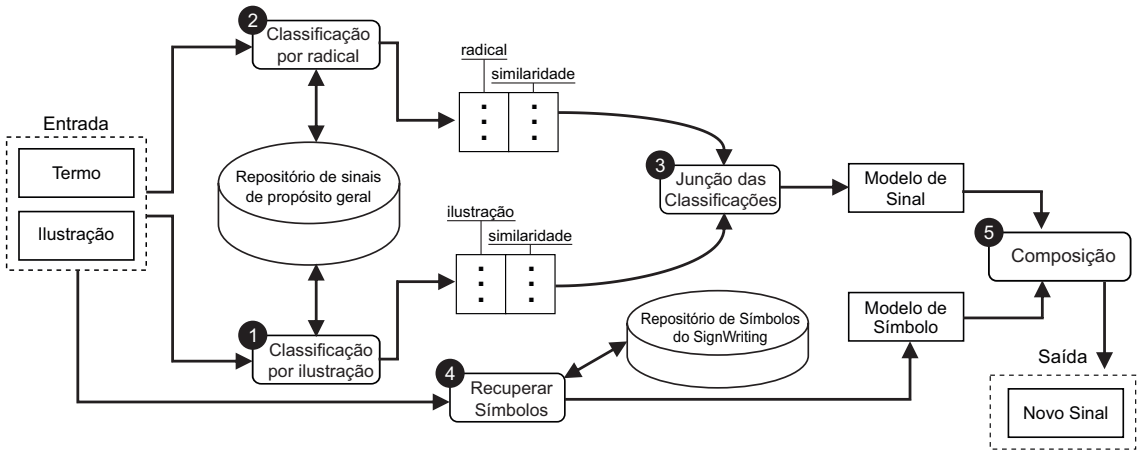
3 Junção das classificações;

4 Recuperação de modelos de símbolo;

5 Composição do novo sinal.

Para a criação de um novo sinal para um conceito de interesse, a abordagem proposta utiliza

Figura 21 – Visão geral do abordagem proposta.

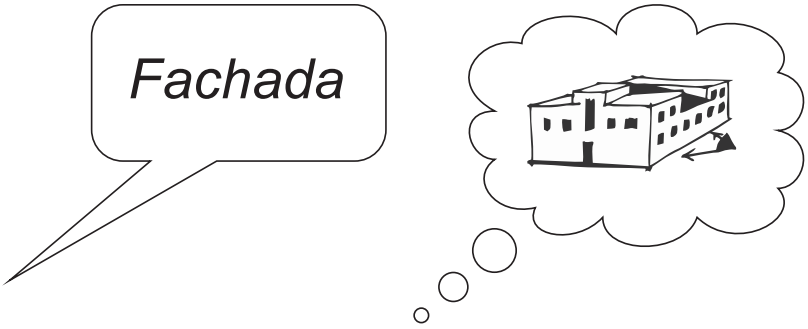


Fonte: Próprio autor.

Figura 22 – Para criação de um novo sinal em Libras para o conceito "fachada", considera-se como entrada o termo "fachada" e a sua ilustração.

(a) Termo do conceito escrito na língua portuguesa.

(b) Ilustração do conceito.



Fonte: Próprio autor.

como entrada dois dados: (1) o termo associado ao conceito na língua oral de referência do surdo e a (2) representação visual ou imagem com ilustração sobre o significado do conceito. A Figura 22 ilustra esses dados utilizando como exemplo o conceito, "fachada" na área do Desenho Arquitetônico.

As seções a seguir descreverão as etapas que compõem a abordagem proposta, a partir da entrada de dados apresentada.

4.1 Classificação por Ilustração

Para recuperar um modelo de sinal apropriado, que será a base do sinal técnico a ser criado para um conceito de interesse (por exemplo, "fachada"), primeiramente compara-se a similaridade do conteúdo visual de uma dada ilustração desse conceito com os conteúdos visuais das ilustrações de termos armazenados em um repositório de sinais de propósito geral.

Basicamente, este repositório é constituído por um amplo conjunto o de sinais bem conhecidos em uma língua de sinais de interesse, acompanhados de seus termos correspondentes na língua oral de referência, bem como de suas respectivas ilustrações. Essas ilustrações contêm uma assinatura visual sobre os significados dos conceitos envolvidos. Cabe ressaltar que assume-se que para cada termo, no repositório, bem como para o conceito de interesse sem sinal, é possível associar uma ilustração.

Especificamente, a abordagem proposta nesta tese descreve o conteúdo visual de cada ilustração por meio de descritores de formas de objetos presentes na ilustração. A forma é uma das características visuais mais importantes de uma imagem e, neste trabalho, é usada para capturar a possível característica icônica associada ao significado do termo.

Com o objetivo de relacionar a forma presente na imagem à iconicidade, buscou-se um descritor de forma robusto e simples, denominado *Shape Context*, proposto por Belongie, Malik e Puzicha (2000). O objetivo quanto a abordar esse descritor é dado por ser um descritor de forma global e que, em certa medida, é robusto à rotação e ao dimensionamento eficiente em reconhecimento de objetos, além de obter bons resultados em vários trabalhos (BHUPTANI; TALATI, 2014; COSTAGLIOLA et al., 2011; JIANG; DREW; LI, 2010). O descritor é gerado por meio do calculo da distribuição relativa das posições dos pontos da forma e, portanto, resumindo em um único descritor global da forma. Posteriormente os descritores *Shape Context* são alinhados e usados para definir uma pontuação para medir a semelhança entre duas formas (BELONGIE; MORI; MALIK, 2006; BELONGIE; MALIK; PUZICHA, 2000). O *Shape Context* é um descritor extremamente rico, sendo insensível à pequena rotação da forma e capaz de capturar o contorno de objetos em uma imagem, conforme apresentado na Subsubseção 2.3.1.1.

O resultado final desta etapa consiste em uma lista de ilustrações associadas aos termos existentes no repositório de sinais de propósito geral, as quais são ordenadas segundo o nível de similaridade de suas informações de forma em relação à informação de forma presente na ilustração do conceito de interesse. Atribui-se a este nível de similaridade de informação de forma valores entre 0 (menos similar) e 1 (mais similar).

4.2 Classificação por Radical

Além da ilustração do conceito de interesse, a abordagem proposta utiliza como informação complementar para criação de um novo sinal, o termo na língua oral de referência do surdo que descreve e/ou nomeia o conceito. Isso se assemelha de forma indireta a seguir o princípio da arbitrariedade, uma vez que está sendo associado o significado do conceito de interesse à uma descrição textual, obtida frequentemente de forma arbitrária.

Assumi-se assim que um novo sinal para um determinado conceito de interesse, pode ser caracterizado por um termo que o qualifica. Neste sentido, será detalhado a seguir o processo de classificação dos sinais modelos que serão candidatos a serem parte do novo sinal a ser criado.

Especificamente, tem-se como objetivo classificar prováveis modelos de sinais a partir de seus termos correspondentes na língua oral de referência. Assim, dado um termo de entrada correspondente ao conceito de interesse, a abordagem proposta retorna uma lista ordenada de modelos de sinais, os quais são ordenados segundo o nível de similaridade dos radicais dos termos referentes a estes modelos em relação ao radical do termo de entrada que designa o conceito de interesse.

O processo de encontrar a raiz de um termo ou palavra é conhecido na computação como *stemming*, que consiste em reduzir as palavras flexionadas para a sua forma raiz ou para o

radical da palavra. Por exemplo, a raiz para o termo "terra" em português é "terr", logo busca-se identificar as palavras com o mesmo radical (possivelmente "terreno", "terreiro", "terrinha" etc.). Para determinação da raiz dos termos da língua oral, utilizou-se o método proposto por Orengo e Huyck (2001) conforme apresentado na Subseção 2.4.1.

Nesse sentido adotou-se a métrica de Levenshtein (LEVENSHTEIN, 1966), para medir a similaridade entre os radicais (sequências de caracteres) de dois termos distintos, e permitir quantificar numericamente a diferença entre tais radicais. Apesar de a métrica inicialmente ser voltada para dados binários, ela foi incorporada em vários trabalhos que precisavam mensurar as diferenças entre duas palavras (KOHONEN, 1985; LYRAS; SGARBAS; FAKOTAKIS, 2007).

Basicamente, consiste em uma função que permite calcular a distância entre duas palavras por meio do número de operações necessárias de inserção, exclusão e substituição de um caractere, com o objetivo de transformar uma palavra em outra. Essencialmente, a ideia baseia-se na hipótese de que quanto menor o número de operações, mais similares são os radicais considerados. Neste contexto, utiliza-se a função de minimização denotada pela Equação 3:

$$D_L(A, B) = \min\{s(j) + i(j) + d(j)\}, \tag{3}$$

em que A e B são os radicais dos termos considerados, s é a operação de substituição, i é a operação de inserção, d é a operação de exclusão e j é o índice que percorre dentro de cada palavra.

Para normalizar o valor da saída da função e aplicá-lo ao cálculo das diferenças entre dois termos, utiliza-se a Equação 4:

$$f_s = 1 - \frac{f_l(t1, t2)}{\max(size(t1), size(t2))}, \tag{4}$$

em que f_s corresponde ao valor de similaridade estimado, f_l denota a função de Levenshtein (LEVENSHTEIN, 1966) em relação aos termos $t1$ e $t2$, e max é o maior tamanho obtido pela função $size$.

Para exemplificar o uso da Equação 4, considere o exemplo apresentado na Figura 23. Para tal exemplo, tem-se que:

$$
\begin{aligned}
f_l &= 5; \\
max(size(s1), size(s1)) &= 13; \\
Similaridade &= \frac{5}{13} \\
Similaridade &= 0,38
\end{aligned}
$$

A função f_l obteve o valor 5, pois é a soma das 3 substituições e de 2 caracteres apagados.

Sendo assim, para cada termo $w \in \mathcal{W}$, em que $\mathcal{W}$ denota um subconjunto de termos com radicais considerados similares, calcula-se a distância por meio da métrica de Levenshtein a fim de definir a distância do termo em relação à língua oral. O resultado do exemplo da Figura 23 é a similaridade de $0,38$ entre os termos $t1$ e $t2$.

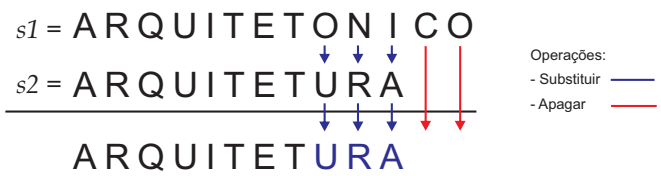
O valor obtido pelo cálculo da distância Levenshtein é normalizado pelo comprimento máximo de termos de entrada. Finalmente, classificam-se os modelos de sinais candidatos de acordo com a distância normalizada de Levenshtein entre o termo do repositório de sinais de propósito geral e o termo referente ao conceito de interesse.

4.3 Fusão das Classificações

A abordagem proposta neste trabalho assume que, para um dado conceito sem sinal, é possível associar um termo na língua oral de referência do surdo, bem como uma ilustração ou representação visual que o descreva. Como mostrado nas seções anteriores, cada uma dessas informações irá gerar uma classificação diferente para possíveis candidatos a sinais modelos. Sendo assim, o objetivo desta etapa consiste em combinar ou fundir essas classificações para recuperar um modelo de sinal que efetivamente represente o conceito de interesse. Acredita-se que, ao considerar as informações em formato textual e visual simultaneamente, exista maior probabilidade de uma recuperação mais adequada de um modelo de sinal semanticamente representativo para o conceito de interesse.

Para a fusão das classificações, primeiro considera-se que os valores retornados de cada uma

Figura 23 – Exemplo das operações da função de Levenshtein.



Fonte: Próprio autor.

das formas de classificação, ou seja, a partir dos termos da língua oral e das ilustrações, são estimativas de quão relevantes os modelos de sinais são para o conceito de interesse. Ao fazer isso, pode-se definir valores limiares para distinguir os níveis aceitáveis ou não de relevância. Em outras palavras, ao estabelecer estes limiares de similaridade para cada classificação, pretende-se evitar modelos de sinais espúrios.

Formalmente, define-se que S^i e ϵ_i indicam, respectivamente, o *ranking* de classificação de similaridade por ilustração e o seu corresponde limiar de similaridade mínimo. Da mesma forma, define-se como S^t e ϵ_t o *ranking* de classificação de similaridade de radicais e seu correspondente limiar de similaridade mínimo. Observa-se que são ignorados os modelos de sinal em S^i e S^t cujos valores de similaridade estão abaixo dos valores de limiares mínimos correspondentes. Deste modo, a partir de S^i e S^t, obtém-se um ranking global de classificação S^m contendo n potenciais modelos de sinais com valores de similaridade s_j^m, para $j = 1, \ldots, n$, os quais são obtidos a partir da Equação 5:

$$s_j^m = (\mathbb{1}[s_j^t > \epsilon_t] \times \mathbb{1}[s_j^i > \epsilon_i]) \times (s_j^t + s_j^i), \tag{5}$$

em que s_j^i e s_j^t representam os valores de similaridade do j-ésimo potencial modelo de sinal no que diz respeito ao conteúdo visual da ilustração e ao radical do termo correspondente, respectivamente, e $\mathbb{1}$ denota a chamada função indicadora, a qual retorna o valor 1 se a expressão lógica usada como argumento for verdadeira e 0, caso contrário. Caso não seja possível selecionar modelos de sinais que satisfaçam simultaneamente os valores de limiares ϵ_t e ϵ_i definidos, simplifica-se a Equação 5 de modo que, neste caso, o valor de similaridade s_j^m do *j*-ésimo potencial modelo de sinal no ranking global S^m seja dado simplesmente por meio da Equação 6:

$$s_j^m = (s_j^t + s_j^i). \tag{6}$$

Sendo assim, por meio do processo de fusão de *rankings* acima descrito, tem-se que o modelo de sinal selecionado corresponderá ao *j*-ésimo sinal associado à primeira posição (maior valor de similaridade) do ranking global S^m obtido.

Na Figura 24 são apresentadas ambas as situações, ou seja, aquela em que se encontra os modelos de sinais com os limiares de similaridade estabelecidos e o outro caso. Consideram-se as classificações de acordo com os termos, *ranking* S^t, e de acordo com as ilustrações, *ranking* S^i. Para cada classificação, há quatro modelos de sinal, isto é, $\{A, B, C, D\}$ e seus respectivos valores de similaridade nos *rankings*. A Figura 24a apresenta pela primeira vez a situação para a qual podem-se encontrar os modelos de sinais que seguem a Equação 5. A classificação já mesclada S^m, conterá apenas o item B, com pontuação $0,9$. Por outro lado,

na 24b, apresenta-se um cenário para o qual atribuí-se uma pontuação zero para todos os modelos de sinais seguindo Equação 5. Observe que os valores de limiar são iguais a $0,3$ e $0,4$, respectivamente. Portanto, aplica-se a Equação 6, em que o resultado final da classificação mesclada é $S^m = \{D, A, B, C\}$.

Observa-se que ao ser retornado apenas o modelo de sinal com maior valor de similaridade global, então, no primeiro caso, tem-se o modelo de sinal B, enquanto que, no último caso, é apresentado o modelo de sinal D.

4.4 Busca pelo símbolo modelo

Um vez obtido o modelo de sinal, realiza-se a seleção de um modelo de símbolo pertencente ao Alfabeto Internacional do SignWriting (SUTTON, 2014), para que então, por meio de um processo de composição de modelos (ver etapa 5 na Figura 21), seja criado o novo sinal desejado.

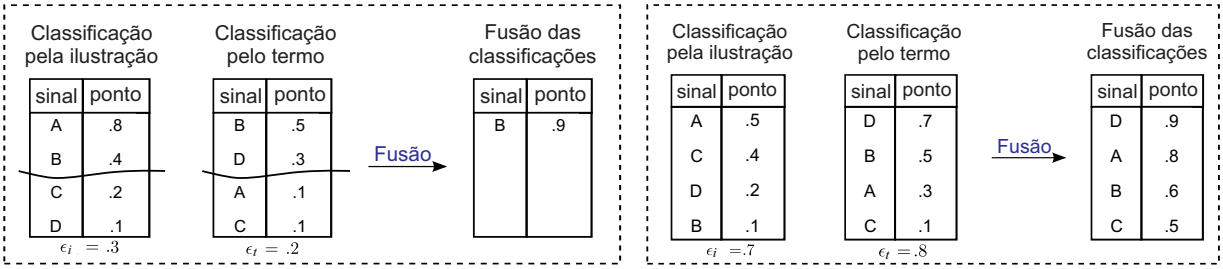
A recuperação ou obtenção do modelo de símbolo (ver etapa 4 na Figura 21) se dá a partir da comparação ou medição de similaridade do descritor de forma associado à ilustração do conceito técnico de interesse com o descritor de forma associado a cada imagem de símbolo existente no respositório de símbolos do SignWriting.

Basicamente, seleciona-se o símbolo que resultar na maior medida de similaridade, isto é, o símbolo cuja informação de forma extraída em sua imagem mais se aproximar da informação de forma da ilustração do conceito de interesse. Novamente, adotou-se o método *Shape Context* para medir a similaridade entre dois descritores semelhantes, o que ocorre na Seção 4.1 por

Figura 24 – Dois exemplos da fusão das classificações, considerando as descrições das imagens e termos. Há quatro sinais, nomeados $\{A, B, C, D\}$, a partir dos quais será selecionado o modelo do sinal.

(a) Dados dois valores de limiar diferentes para cada classificação, ou seja, $\epsilon_i = 0,3$ para classificação por imagem; e $\epsilon_t = 0.2$, para classificação por termo. A classificação final contém apenas o sinal B.

(b) Dado $\epsilon_i = 0,7$, para classificar por imagem; e $\epsilon_t = 0,8$, para classificar pelo termo. Observa-se que nenhum sinal atinge os limiares. A classificação final contém todos os sinais. O modelo de sinal é D.



Fonte: Próprio autor.

meio de uma função de similaridade.

4.5 Composição do novo sinal

Nesta etapa, já se tem definido o modelo de sinal e o modelo de símbolo para o novo sinal que irá representar o conceito de interesse na língua de sinais. De posse desses elementos, será necessário combinar esses modelos selecionados para produzir um novo sinal.

Alguns trabalhos relatam a importância de um símbolo perante o sinal, e fazem com que a inclusão de um novo símbolo em um sinal já existente permite transformá-lo em outro sinal, com sua própria identidade (BARRETO, Madson; BARRETO, 2015; STIEHL et al., 2015; AERTS et al., 2004) Para auxiliar na composição deste novo sinal e permanecer dentro dos padrões aceitáveis, seguiram-se as principais regras do sistema SignWriting para escrever um sinal (SUTTON, 2014; BARRETO, Madson; BARRETO, 2015).

Uma característica-chave de SignWriting é o uso de um "SingBox" invisível, que simplesmente se refere a um *layout* bidimensional de símbolos (THIESSEN, 2011). Assim, dado um conceito que está sendo representado, utiliza-se as posições relativas dos símbolos dentro do SingBox para representar iconicamente as posições das partes do corpo, como as mãos. Note que diferente da relação de uma sequência de caracteres dentro de uma palavra nas línguas faladas, em SignWriting, não existe uma relação linear óbvia entre os símbolos dentro do SingBox.

No entanto, fora do SingBox do sinal, o sistema SignWriting considera uma relação linear. Visualmente, os sinais são frequentemente escritos em colunas verticais, ou seja, o SingBox é organizado de cima para baixo dentro da coluna, conforme apresentado na Subsubseção 2.1.4.1. Uma vez que se tem o objetivo de combinar um sinal e um símbolo, considera-se o posterior como um sinal parcial. Portanto, segue-se a relação linear entre eles para representar visualmente sua combinação e produzir um novo sinal. Consideras-se, assim, que esse arranjo é uma boa combinação entre a simplicidade, a estética, a iconicidade e as principais regras do sistema SignWriting.

Capítulo 5

Resultados experimentais

Neste capítulo são apresentados os detalhes relacionados aos experimentos realizados para validação da presente proposta, valendo-se dos seguintes aspectos: (i) avaliação da abordagem, (ii) conjunto de sinais a serem utilizados como referência, (iii) metodologia de validação da proposta, e (iv) referências aos conjuntos de sinais e implementações utilizadas. São também apresentados e discutidos os resultados alcançados sobre a ótica da análise quantitativa e qualitativa.

5.1 Ambiente de experimentos

Uma metodologia para avaliação dos sinais criados é algo extremamente complexo. A abordagem presente neste trabalho, faz-se necessário elucidar, tem como proposta auxiliar na criação de sinais terminológicos para a comunidade surda. Logo, faz-se necessária a avaliação desses sinais pelos membros desta comunidade a fim de validá-los. Além disso, os sinais também devem ser analisados por especialistas que possuam conhecimento tanto na língua de sinais quanto na área técnica para a qual foram gerados.

Para a validação deste trabalho de neologismos terminológicos para língua de sinais, se fez inevitável delimitar uma língua de sinais e a área técnica para a qual seriam criados os sinais. Sendo assim, as escolhas foram a Libras e a área do desenho arquitetônico. Para seleção da língua, foram considerados os seguintes aspectos: o país no qual estamos, o auxílio de surdos nativos e intérpretes, um dicionário geral ilustrado (no caso CAPOVILLA), além de permitir colaborar com a desenvolvimento da língua de sinais no país. Quanto à área para a geração dos sinais, a escolha se deu pelo acesso a um conjunto de sinais que foram criados de modo convencional, e por possuirmos pessoas com conhecimento técnico na área e na língua de sinais.

Foram selecionadas dezessete pessoas que atendiam à especificação acima mencionadas e que autorizaram o uso dos dados, formando, assim, um grupo de avaliadores dos sinais. Os

indivíduos que participaram do grupo de avaliação são compostos por 35% de surdos nativos e 64% de intérpretes de Libras. Quanto ao conhecimento da Libras, o grupo é considerado homogêneo.

Acredita-se que este é um grupo representativo, que permite avaliar de maneira justa nossa abordagem de criação de sinais técnicos. Embora no Brasil a transcrição dos sinais de Libras em SignWriting esteja em vigor desde 1997 com o SignNet Project (COSTA, 2001) e os dicionários de Capovilla (CAPOVILLA; RAPHAEL, 2001a; CAPOVILLA; RAPHAEL, 2001b), observa-se que apenas um pequeno grupo de pessoas é realmente fluente na escrita em SignWriting e, geralmente, são usuários especializados.

O processo de avaliação dos sinais pelos colaboradores não poderia ser extenso (UNION, 1998), logo, limitou-se à avaliação de 10 conceitos técnicos da área de desenho arquitetônico com base no trabalho de Lima (2014). Acredita-se que este seja um número suficiente para avaliar a qualidade da abordagem proposta. Conforme já discutido no Capítulo 3, é importante ressaltar que o abordagem aqui proposta é completamente inédita, uma vez que, não existe metodologia semelhante na literatura, que trate da criação de sinais assistida por computador e que possa ser usada como uma referência adicional a este trabalho.

Os sinais criados manualmente e propostos em Lima (2014) foram confrontados com os sinais gerados automaticamente pela abordagem proposta para os mesmos termos, sendo conduzindo um método de avaliação de comparação intercalada.

Cada sinal é exibido aos membros do nosso comitê de avaliação dos sinais gerados da seguinte forma: é apresentada uma descrição textual em português, seguida de uma descrição em Libras (por meio de um vídeo), pelo termo oral do conceito de interesse para o qual queremos criar um novo sinal (por exemplo, bússola) e, por último, por uma ilustração sobre o conceito de destino.

Nas repostas dos avaliadores, foi considerada uma variedade de pontuações com a finalidade de saber a relevância do sinal perante as características avaliadas com o propósito de medir a qualidade dos sinais criados. Na literatura, esse tipo de abordagem para avaliação é conhecido e tem sido utilizado com sucesso em diferentes áreas (BAEZA-YATES; RIBEIRO-NETO, 2011; CHAPELLE et al., 2012; CHUKLIN et al., 2015). A proposta é confrontar os sinais, o manual e o gerado pela presente abordagem, em um par de classificações (*rankings*).

Logo os avaliadores classificavam os sinais apresentados de acordo com sua opinião, sendo assim a interação dos avaliadores com os votos produziam um classificação dos sinais avaliados, permitindo assim a preferência. Acredita-se que esta metodologia de avaliação garante uma avaliação imparcial da presente abordagem.

É importante enfatizar que, pela natureza deste trabalho ter o envolvimento de humanos no

processo de avaliação dos sinais, fez-se necessária a submissão de um projeto ao Comitê de Ética em Pesquisa (CEP), que é um colegiado multidisciplinar e transdisciplinar independente, para o qual toda instituição que realiza pesquisa envolvendo seres humanos no Brasil submete seus trabalhos, com a finalidade de defender os interesses dos sujeitos da pesquisa em sua integridade e dignidade, contribuindo para o desenvolvimento da pesquisa dentro de padrões éticos. Com isso, foi obtido o parecer favorável de número 1.635.736, em julho de 2016, conforme o Apêndice A.

Quanto ao procedimento de coleta de dados, este trabalho considerou uma abordagem subjetiva baseada na percepção de observadores humanos. Os dados foram coletados em ambientes naturais aos avaliadores, especificamente, nas casas ou nos escritórios dos envolvidos, onde cada participante, usando um dispositivo pessoal com acesso à internet, respondeu a um formulário com o objetivo de avaliar a qualidade do sinais, que foram classificados de acordo com uma escala de cinco pontos. Cada participante teve no máximo 30 minutos para responder o formulário de avaliação.

Nesta abordagem subjetiva, usou-se uma única estratégia de estímulo, ou seja, os sinais criados pelo nosso método e aqueles criados de forma manual foram exibidos um a um, e avaliados individualmente por cada participante. No entanto, esses não eram apresentados de forma sequencial e o participante não sabia a origem da criação do sinal.

Essa metodologia é frequentemente utilizada pelo *Video Quality Experts Group*, que em livre tradução seria Grupo de Especialistas em Qualidade de Vídeo, a fim de avaliar qualidade de conteúdo de vídeo (UNION, 1998; ROHALY et al., 2000). Essa abordagem é denominada Classificação Absoluta com Referência Oculta. Em inglês o termo é conhecido com *Absolute Category Rating with Hidden Reference* (ACR-HR), e foi aplicado no experimento de avaliação com a intenção de garantir uma comparação justa (ITU-T, 2008).

5.1.1 Repositório de Dados

Para o desenvolvimento deste trabalho, foi adotado um conjunto de repositório de dados. Primeiro, temos um dicionário de propósito geral de sinais (CAPOVILLA; RAPHAEL, 2001a; CAPOVILLA; RAPHAEL, 2001b), que consiste em mais de 3.000 sinais contendo três dados importantes que são:

i. termo em português;

ii. ilustração do conceito;

iii. sinal em Libras.

O segundo conjunto repositório de dados é o dicionário conhecido como **SignBank**, disponível em <http://www.signbank.org>, o qual consiste em sinais na notação escrita SignWriting,

com mais de 7.050 sinais. O dicionário é especialmente útil, pois contém sinais representados em termos de combinações de símbolos, sendo utilizado posteriormente para selecionar um modelo de símbolo que seja combinado com o sinal modelo para produzir o novo sinal.

E por fim, o conjunto de termos técnicos produzidos por Lima (2014). Esse conjunto de dados possui mais de 70 sinais da área do desenho arquitetônico, como mencionado anteriormente, e teve como finalidade comparar esses sinais criados manualmente com a abordagem proposta por este trabalho. Desse conjunto foram selecionados 10 termos, com base em duas características de formação em português:

1. Imagem: para a geração do sinal é necessário que o termo possua uma imagem representativa. A partir dessa necessidade, foi selecionada uma imagem para cada termo no Dicionário Ilustrado de Arquitetura de Albernaz e Lima (1998);
2. Formação: considerando os termos que possuíam imagens, foram avaliados seus aspectos de formação (descritos na Subseção 3.3.4) e, foram selecionados pelo menos um sinal para cada aspecto.

Esses termos selecionados para validação são apresentados na Tabela 5, na qual também são informados os métodos de formação dos mesmos, no âmbito do trabalho de Lima (2014).

5.1.2 Questionário

A partir dos sinais desenvolvidos pela presente metodologia e os criados manualmente, foi elaborado um questionário que buscasse obter do grupo avaliador informações quanto à qualidade e conhecimento acerca dos sinais. Para isso, o formulário foi elaborado com a participação de membros da comunidade acadêmica, bem como de membros da comunidade surda os quais fazem parte do Grupo de Pesquisas Interdisciplinares em Informação Multimídia do CEFET-MG.

Tabela 5 – Tabela de termos e seus métodos de formação.

Termo	**Método de formação**
compasso	empréstimo linguístico, iconicidade
corrimão	justaposição ou composição, iconicidade
cortes	composição
fachada	composição
maçaneta	empréstimo linguístico, iconicidade
patamar	justaposição ou composição, composição
pilar	iconicidade
platibanda	justaposição ou composição, composição, iconicidade
porão	justaposição ou composição
rodapé	iconicidade

O questionário possui uma primeira parte, que visou à identificação do entrevistado, por meio das perguntas de 1 a 13. Posteriormente, buscou-se obter para cada sinal as informações referentes ao caráter morfológico e à sua qualidade, que são as perguntas de 14 a 25. As perguntas presentes no questionário foram:

1) Nome Completo?
2) Identificação (RG ou CPF)?
3) Data?
4) Aceito a participação na pesquisa?
5) E-mail?
6) Escolaridade?
7) Qual a formação / atuação?
8) Data de nascimento?
9) Surdo?
10) Com qual idade começou a falar Libras?
11) Conhece o sistema de escrita SignWriting?
12) Quanto tempo conhece o SignWriting?
13) Qual a sua familiaridade com o SignWriting
14) Conhecia o conceito antes?
15) Conhecia o sinal antes?
16) É possível sinalizar o sinal?
17) Nota do conhecimento dos símbolos usados no sinal?
18) Na sua opinião o sinal é do tipo?
19) É feito com uma ou com duas mãos?
20) Algum tipo de contato?
21) Qual o plano da seta de movimento?
22) Possui expressões não manuais?
23) Nível de satisfação com o sinal?
24) O nível de associação do conceito do termo ao sinal?
25) Sugestão para quanto ao sinal?

Para as perguntas de 23 a 24, o avaliador tem a opção de selecionar um valor para o sinal, conforme apresentado na Tabela 6, sendo 1 para classificá-lo como "péssimo", e 5 para classificá-lo como "excelente".

Tabela 6 – Análise dos processo de formação de Sinais.

Escala	Qualidade
5	Excelente
4	Bom
3	Aceitável
2	Ruim
1	Péssimo

Fonte: (MAIA, 2015)

As questões morfológicas, tratadas pelas perguntas de 18 a 22, contribuem para o preenchimento de alguns itens da ficha lexicográfica, os quais serão armazenados junto ao sinal no repositório. Essas características fazem parte do aspecto *Usual* do sistema. Para avaliar o aspecto científico, são utilizadas as perguntas 23 e 24, que buscam obter informações sobre a qualidade do sinal.

Para responder o questionário e com isso avaliar e validar os sinais, foi selecionado um grupo seleto de usuários, que são integrantes da comunidade surda com notório conhecimento na escrita de sinais. A aplicação do questionário ocorreu por meio de uma plataforma disponibilizada na Internet.

5.1.3 Métricas

Neste trabalho, foram adotadas duas métricas para avaliar o desempenho da presente abordagem. Primeiramente, foi adotada a métrica Opinião Média das Avaliações, que é adotada na literatura inglesa, e é denominada como *Mean Opinion Score* (MOS). O MOS é uma métrica com valores inteiros entre 1 e 5, conforme apresentado pela Tabela 6. Note-se que os valores mais elevados do MOS significam maior qualidade no sinal avaliado. Em outras palavras, MOS estima a percepção humana da qualidade do sinal.

A segunda métrica utilizada por este trabalho é a Diferença das Médias das Avaliações, também muito comum na literatura inglesa e denominada como *Diference Mean Opinion Score* (DMOS). O DMOS é uma métrica associada a cada sinal e representa sua qualidade subjetiva. Em certa média, o objetivo desta métrica é apresentar a diferença entre a abordagem de referência, no caso desenvolvido de forma manual, e a abordagem proposta, podendo variar de 0 a 4.

Cabe ressaltar que tal abordagem das métricas de avaliação representa uma proposta inédita na literatura, na qual se evidencia uma forte carência por metodologias científicas que auxiliem no estudo e na determinação do nível de qualidade de sinais gerados no âmbito de uma determinada língua.

5.1.4 Reprodutibilidade

Quanto à reprodutibilidade, os conjuntos de dados utilizados nos experimentos estão disponíveis gratuitamente.

O algoritmo de radicalização adotado é de licença livre e disponível em: <https://code.google.com/archive/p/ptstemmer>. A implementação do descritor de imagem abordado por este trabalho, Shape Context de Belongie, Malik e Puzicha (2000) está disponível em: <https://goo.gl/dxUgJ8>. Quanto aos sinais em SignWriting, a base de dados SignBank é pública e disponível em: <http://www.signbank.org>. Portanto, é possível a reprodutibilidade dos experimentos, que podem ser usados para melhorar a abordagem proposta e para outras linhas futuras de pesquisa.

5.2 Discussões

Nesta seção, serão discutidos os resultados obtidos pela abordagem proposta por este trabalho. Como dito anteriormente, ressalta-se que não há na literatura métodos assistidos por computador para criar sinais para conceitos técnicos ou não técnicos. Sendo assim, comparamos a presente abordagem com um dicionário de conceitos técnicos que foi desenvolvido com a experiência em linguagem de sinais na área de desenho arquitetônico. Convém destacar que o objetivo não é propor sinais que sejam melhores do que os criados de forma manual.

O principal objetivo deste trabalho é ajudar na criação de novos sinais, propondo sinais tão precisos quanto possível e aceitos pela comunidade surda. Observa-se que os sinais propostos pela abordagem deste trabalho, podem atuar como ponto de partida na criação de novos sinais, que precisam ser revisados por uma comissão qualificada para produzir termos em língua de sinais que sejam aceitáveis. Espera-se que um grande conjunto de sinais criados ajude na construção de forma escalável de dicionários técnicos em linguagens de sinais. Doravante, haverá referência à criação dos sinais de forma manual e à abordagem proposta para criação de sinais assistida por computador, como CMS e CSAC, respectivamente.

5.2.1 Satisfação da proposta

Quanto à satisfação dos sinais propostos pela abordagem, os avaliadores classificavam os sinais seguindo os valores do MOS, sendo que valores maiores são melhores, de tal forma que o valor 1 é classificado como "péssimo" e o valor 5 classificado como "excelente". Especificamente, os sujeitos precisam responder sobre a satisfação com relação aos sinais apresentados.

Os resultados da média das avaliações quanto ao nível de satisfação dos sinais é apresentado na primeira linha da Tabela 7 e na Figura 25. Observa-se que, em média, o método CMS atinge um valor MOS de 3,3765. Isso demonstra que não há uma unanimidade entre os assuntos sobre

Tabela 7 – Resumo dos resultados das métricas.

	Média MOS		
Critério	CMS	CSAC	Média DMOS
Nível de Satisfação	3,7565	2,4529	0,9235
Precisão Semântica	3,3235	2,4824	0,8412

o melhor sinal para representar um conceito de interesse. Enquanto isso, a presente proposta, CSAC, atinge um valor do índice MOS de 2,4529.

O teste de hipótese estatístico definido para esse trabalho é:

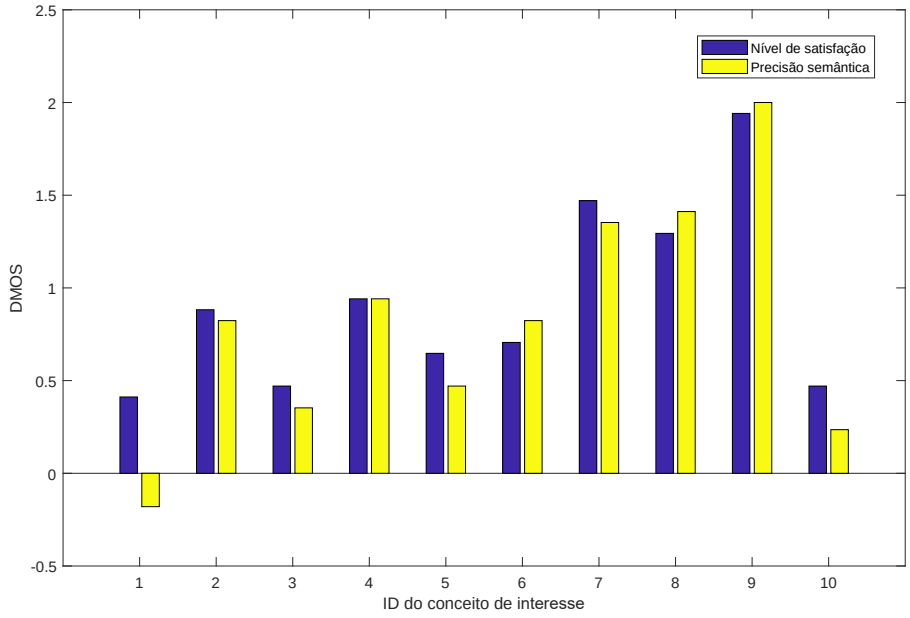
$$H_0 : CMS = CSAC, H_1 : CMS \neq CSAC, \tag{7}$$

sendo que a hipótese nula $H_0 : CMS = CSAC$ considera que não há diferença significativas nas avaliações entre as abordagens para criação de sinais, enquanto o a hipótese alternativa $H_1 : CMS \neq CSAC$, considera que há diferença significativas entre as avaliações das abordagens para criação de sinais.

Uma vez que nossos dados não são normalmente distribuídos, ou seja, não é possível determinar uma distribuição, torna-se necessária a escolha de um teste não paramétrico para amostras pareadas, como teste de hipóteses. Logo, foi adotado dois testes estatísticos, o Teste de Sinais (TS) e o Teste de Wilcoxon (TW), considerando como margem significativa de $p < 0,05$.

Conforme a Tabela 8, o TW demonstrou que há diferenças significativas entre as abordagens

Figura 25 – Desempenho da abordagem proposta da métrica DMOS para cada conceito de interesse.



Fonte: Próprio autor.

Tabela 8 – Resultado dos testes: Wilcoxon (TW) e Sinais (TS). N = número de amostras, Z valor da estatística do teste e p é probabilidade associada. Os valores em negrito mostram diferenças estatisticamente significativas.

Termo	Questão	Teste de Wilcoxon (TW)			Teste de Sinais (TS)		
		N	Z	p	N	Z	p
1	23	11	1,2448	0,2132	11	0,6030	0,5465
	24	10	0,0000	1,0000	10	-0,3162	0,7518
2	23	13	1,6074	0,1080	13	1,1094	0,2673
	24	12	1,5297	0,1261	12	1,4434	0,1489
3	23	11	0,9780	0,3281	11	0,0000	1,0000
	24	10	0,7135	0,4755	10	-0,3162	0,7518
4	23	9	2,0732	**0,0382**	9	1,3333	0,1824
	24	10	2,0896	**0,0367**	10	1,5811	0,1138
5	23	8	2,1004	**0,0357**	8	1,7678	0,0771
	24	9	1,3624	0,1731	9	1,3333	0,1824
6	23	12	2,0004	**0,0455**	12	1,4434	0,1489
	24	9	2,1917	**0,0284**	9	1,3333	0,1824
7	23	13	2,9003	**0,0037**	13	2,2188	**0,0265**
	24	13	2,6207	**0,0088**	13	2,2188	**0,0265**
8	23	14	2,1972	**0,0280**	14	1,3363	0,1814
	24	14	2,4169	**0,0157**	14	1,3363	0,1814
9	23	14	3,2958	**0,0010**	14	3,4744	**0,0005**
	24	14	3,2958	**0,0010**	14	3,4744	**0,0005**
10	23	7	1,0142	0,3105	7	0,7559	0,4497
	24	8	0,5601	0,5754	8	1,0607	0,2888

CMS e CSAC para os termos: 4, 5 (apenas o nível de satisfação), 6, 7, 8 e 9. Já o TS mostrou diferenças significativas nos termos 7 e 9. Apesar da diferença entre observada entre os dois testes estatísticos estes resultados reforçam que os termos 7 e 9 não foram bem aceitos pelos avaliadores, conforme já demostrado pelo Figura 25. Além disso o TW reforça, para o termo 8, o resultado apontado pela métrica DMOS, sugerindo uma não aceitação deste termo pelos avaliadores. Conclui-se que a abordagem apresentada, que é assistida por computador, possui índice de equivalência aceitável comparado ao método manual proposto em Lima (2014).

A Figura 25 expõe no gráfico a métrica DMOS em relação aos conceitos de interesse. Observa-se que, para conceitos como 1 e 10, a diferença entre CMS e CSAC é pequena. No primeiro caso, temos valores de MOS para CMS e CSAC igual a 4,6 e 3,8 respectivamente. Acredita-se que é um conceito relativamente fácil uma vez que a abordagem manual e a abordagem proposta alcançam um alto acordo entre as avaliações. Por outro lado, no segundo conceito de interesse (ou seja, o 10°), os valores de MOS são 1,8 e 1,2, o que leva a uma pequena diferença como antes. Assim, conclui-se que alguns conceitos de interesse são difíceis tanto para o manual quanto para as abordagens auxiliadas por computador.

5.2.2 Associação entre sinal e significado

Esta subseção tem como objetivo avaliar o grau de associação entre o sinal e o real significado do conceito de interesse. Os avaliadores devem considerar a qualidade e as características morfológicas do sinal. Os resultados obtidos com este propósito de investigação, que é a associação entre sinal e significado, é demostrado na segunda linha de Tabela 7 e em Figura 25.

Observa-se, primeiramente, que na Tabela 7 há os resultados comparáveis quanto ao desempenho da abordagem automatizada e o método manual. Tem-se um valor MOS de 2,4824 para CSAC, enquanto o método CMS atinge 3,3235. Considerando o resultado obtido, argumenta-se que seja promissor, pois pode-se assumir que é possível obter sinais tão eficazes quanto os propostos por uma abordagem manual.

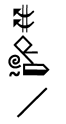
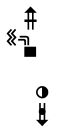
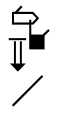
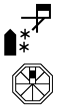
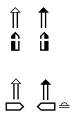
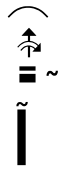
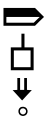
Na Figura 25 apresentam-se os valores do DMOS para as abordagens investigadas. Nota-se, primeiramente, que se tem um pequeno valor de DMOS, ou seja, de 0,2, ao considerar o primeiro conceito de interesse. Em outras palavras, há uma pequena diferença entre a taxa dada pelos especialistas para a CMS e a abordagem CSAC. Conclui-se então que o sinal proposto é tão efetivo quanto o sinal construído manualmente por especialistas.

5.2.3 Analise qualitativa

O foco desta subseção consiste em apresentar os sinais gerados pelas abordagens do CMS e do CSAC. Na Tabela 9, são apresentados os sinais gerados por cada abordagem para cada conceito de interesse. Nota-se que a ordem dos sinais é a mesma que a apresentada na Figura 25. Sabe-se que é extremamente difícil avaliar a qualidade e o nível de associação do significado, no entanto existem alguns pontos que devem ser apresentados.

Observa-se, primeiramente, que o sinal proposto pela abordagem automatizada para o conceito ID 1, que se refere a "compasso", é muito semelhante ao sinal criado manualmente. Pode-se concluir também que, mesmo quando os sinais são muito distintos entre si, por exemplo, o conceito de ID 10, ambos os sinais são aceitáveis pelo comitê de avaliação.

Tabela 9 – Os dez sinais propostos pelas abordagens CMS e CSAC para cada conceito técnico do desenho arquitetônico.

Termo	Ilustração	Projeto de Lima	Sistema Proposto
compasso			
corrimão			
cortes			
fachada			
maçaneta			
patamar			
pilar			
platibanda			
porão			
rodapé			

Capítulo 6

Conclusão

O presente trabalho permite subsidiar o desenvolvimento de neologismos terminológicos em língua de sinais em CTEM, por meio de uma abordagem inovadora que utiliza como base teorias da computação e da linguística, contribuindo no processo de aprendizagem de estudantes surdos.

Ao adotar, na abordagem, propriedade visual do conceito de interesse para a criação do novo sinal, teve como princípio relacionar o conceito de iconicidade por meio de uma ilustração. Em certa medida, a proposta adotou a teoria da computação, em específico da visão computacional, para analisar a ilustração, definindo características que determinam a sua forma. Além da propriedade visual, adotaram-se estratégias já conhecidas para recuperação de textos, relacionadas à programação da linguagem natural. Essas propriedades em conjunto visavam buscar sinais modelos, que, em certa medida, assemelhavam-se ao conceito de interesse.

Diferente da grande maioria dos trabalhos da literatura, os quais adotam a tecnologia como ferramenta complementar à aprendizagem, a presente tese coloca a tecnologia como aliada neste processo e como um facilitador neste desenvolvimento. Ao que tudo indica, pelo estado da arte, este é o primeiro trabalho que adota metodologias computacionais para automatizar o processo de construção de novos sinais. Esta criação de sinais permitirá o desenvolvimento de um novo léxico técnico, trazendo para a comunidade surda um grande benefício, auxiliando-a no aprendizado e na compreensão de novos conceitos técnicos de forma rápida e bem definida.

Por se tratar de um trabalho inovador, foi necessária a comparação de desempenho dessa abordagem com a abordagem clássica, que é a construção manual de sinais. Por meio de experimentos empíricos com especialistas humanos, mostrou-se que a presente abordagem atinge desempenho comparável. A taxa de aprovação que a presente abordagem alcançou foi de 60% dos sinais desenvolvidos. Ao que tudo indica, esse trabalho será um marco importante no processo de desenvolvimento do léxico e auxiliará na construção de novos sinais para a língua de sinais, trazendo, assim, grandes benefícios para a comunidade surda.

6.1 Resultados alcançados

Durante o desenvolvimento da presente Tese foram alcançados alguns resultados que foram divulgados com a publicação em dois periódicos:

- SOUZA, C. L.; PÁDUA, F. L. C.; LIMA, V. L. S.; LACERDA, A.; CARNEIRO, C. A. G. A computational approach to support the creation of terminological neologisms in sign languages. *Computer Applications in Engineering Education*, v. 26, n. 1, p. 1–14, 2018. Disponível em: <http://doi.wiley.com/10.1002/cae.21904>. Com classificação Qualis da CAPES B1 nas áreas Interdisciplinar e Ciência da Computação;
- SOUZA, C. L. DE; LIMA, V. L. S. E; PÁDUA, F. L. C. Abordagem interdisciplinar para a criação e preservação de novos sinais para dicionários terminológicos em libras. *Acta Semiótica et Lingvistica*, v. 19, n. 1, p. 76–90, 2014. Com classificação Qualis da CAPES B2 na área Interdisciplinar.

Além dos periódicos, há um pedido de depósito de patente sob o processo número **BR 10 2017 002163 7**, tendo como depositárias as instituições: CEFET-MG, IF Sudeste MG e FAPEMIG junto ao Instituto Nacional da Propriedade Industrial (INPI).

E por fim, o trabalho foi apresentado em dois eventos que tinham como foco o desenvolvimento da língua de sinais, sendo assim, permitiram o reconhecimento do projeto perante a comunidade surda:

- I Encontro Nacional de Ensino e Letramento em Escrita da Língua de Sinais (ENELELS) promovido pela Universidade Federal do Ceará em 2016;
- I Congresso Nacional de Pesquisas em Linguística e Línguas de Sinais promovido pela Universidade Federal de Santa Catarina em 2016.

6.2 Trabalhos futuros

O mecanismo de criação de sinais apresentado por este trabalho utilizou-se de técnicas robustas e eficazes. Sendo assim, poderia expandir o conjunto de conceitos de interesse para a área do desenho arquitetônico que teria seus sinais correspondentes. Além disso, seria possível também o desenvolvimento de novos conjuntos de sinais para outras áreas do CTEM, as quais estão além do domínio da área desenvolvida por este trabalho.

Este trabalho utilizou a Libras para testar sua abordagem. Sendo assim, cabe validar a proposta em outro idioma, por exemplo, a *American Sign Language*, permitindo, assim, a criação de parcerias com outras instituições de ensino e de pesquisa.

As técnicas computacionais adotadas nesta tese mostram-se eficazes, porém é possível pensar em aprimoramentos com novas metodologias, que abordariam em certa medida a relação de ilustração do sinal sob a ótica da iconicidade. É possível vislumbrar a utilização de técnicas computacionais com o aprendizado de máquinas, como a aprendizagem profunda (*Deep Learning*), que busca modelar abstrações de alto nível de dados, como nas imagens, por meio de um grafo com várias descrições de transformações lineares e não lineares associadas a sinais existes. Esse tipo de aprendizagem permitirá o reconhecimento de padrões, aprimorando, assim, o mecanismo de geração de novos sinais de forma mais independente. Nesse caso, a utilização do sinal modelo, como foi proposta por esse trabalho, não seria necessária.

Além do aprimoramento do gerador de sinais, é de extrema importância investir em uma proposta que permita consolidar os sinais por meio de uma plataforma digital, o que seria um grande repositório dos sinais gerados e validados para a divulgação do conhecimento técnico na comunidade. Isso poderia resultar em uma organização que fomentaria recursos terminológicos para a língua de sinais.

Referências

AARONS, D.; AKACH, P. South african sign language - one language or many? a sociolinguistic question. **Stellenbosch Papers in Linguistics**, v. 31, p. 1–28, 1998. ISSN 2223-9936. Disponível em: <http://spil.journals.ac.za/pub/article/view/55>. Citado na página 13.

AERTS, S. et al. Searching signwriting signs. **4th International Conference on Language Resources and Evaluation (LREC 2004), Sign Processing workshop, Lisbon, Portugal**, p. 79–81, 2004. Citado na página 57.

ALBERNAZ, M. P.; LIMA, C. M. **Dicionário Ilustrado de Arquitetura**. 1. ed. São Paulo, SP: ProEditores, 1998. 316 p. Citado na página 61.

ARROTÉIA, J. Dissertação (Mestrado em Linguística), **O papel da marcação não-manual nas sentenças negativas em Língua de Sinais Brasileira (LSB)**. Campinas, SP: [s.n.], 2005. 131 p. Citado na página 14.

BAEZA-YATES, R.; RIBEIRO-NETO, B. **Modern Information Retrieval: The Concepts and Technology behind Search**. 2. ed. New York, NY: Addison-Wesley Professional, 2011. v. 1. 944 p. ISBN 978-0321416919. Citado na página 59.

BARRETO, Madson; BARRETO, R. **Escrita de sinais sem mistérios**. 2. ed. Salvador, BA: Libras Escrita, 2015. 416 p. ISBN 978-85-69197-00-3. Citado na página 57.

BARTH, C.; SCHNEIDER, F. C.; SILVA, A. A. Construção da leitura/escrita em língua de sinais de crianças surdas em ambientes digitais. **Revista Novas Tecnologias na Educação - RENOTE**, v. 7, n. 3, p. 11, 2009. ISSN 1679-1916. Citado 3 vezes nas páginas 35, 40 e 41.

BATTISON, R. M. Phonological deletion in american sign language robbin battison sign. **Sign Language Studies**, v. 1005, n. 1, p. 1–19, 1974. ISSN 1533-6263. Disponível em: <http://muse.jhu.edu/content/crossref/journals/sign{_}language{_}studies/v1005/5.battiso>. Citado 2 vezes nas páginas 15 e 16.

BATTISON, R. M. **Lexical borrowing in American sign language**. Burtonsville: Linstok Pr, 1978. 240 p. ISBN 978-0932130020. Citado na página 15.

BELLUGI, U.; FISCHER, S. A comparison of signed and spoken language. **Cognition**, v. 1, n. 2-3, p. 173–200, 1972. ISSN 00100277. Citado na página 13.

BELONGIE, S.; MALIK, J.; PUZICHA, J. Shape context: A new descriptor for shape matching and object recognition. In: LEEN, T. G. D. T. K.; TRESP, V. (Ed.). **Advances in Neural Information Processing Systems 13: Proceedings of the 2000 Conference**. Cambridge: MIT Press, 2000. p. 831–837. ISSN 0162-8828. Disponível em: <dx.doi.org/10.1109/34.993558>. Citado 4 vezes nas páginas 30, 31, 52 e 64.

BELONGIE, S.; MALIK, J.; PUZICHA, J. Shape context: A new descriptor for shape matching and object recognition. In: LEEN, T. G. D. T. K.; TRESP, V. (Ed.). **Advances in neural information processing systems**. Cambridge: MIT Press, 2001. p. 831–837. ISSN 0162-8828. Disponível em: <http://dx.doi.org/10.1109/34.993558>. Citado na página 30.

BELONGIE, S.; MORI, G.; MALIK, J. Matching with shape context. In: KRIM, H. (Ed.). **Statistics and Analysis of Shapes**. Boston, MA: Springer, 2006. p. 81–105. ISBN 978-0-8176-4481-9. Citado na página 52.

BERNARDINO. Dissertação (Mestrado em Linguística), **A construção da referência por surdos na LIBRAS e no português escrito : a lógica no absurdo**. Belo Horizonte, MG: [s.n.], 1999. Citado 2 vezes nas páginas 40 e 44.

BHUPTANI, N.; TALATI, B. Variations in shape context descriptor: A survey. **Internation Journal of Computer Applications**, v. 90, n. 12, p. 29–33, 2014. ISSN 0975-8887. Citado na página 52.

BIBER, D. et al. **Longman grammar of spoken and written English**. London: MIT Press Cambridge, 1999. 1203 p. ISBN 978-0582237254. Citado 2 vezes nas páginas 12 e 13.

BIGHAM, J. P. et al. Asl-stem forum: A bottom-up approach to enabling american sign language to grow in stem fields. In: **Second International Conference on Weblogs and Social Media (ICWAW)**. Seattle, Washington, USA: [s.n.], 2008. p. 176–177. ISBN 9781577353553. Citado 3 vezes nas páginas 38, 47 e 48.

BRASIL. **Lei Nº 10.436, de 24 de abril 2002**. 2002. Disponível em: <http://www.planalto.gov.br/ccivil{_}03/leis/2002/L10436.> Citado na página 26.

BRASIL. **Decreto Nº 5.626, de 22 de dezembro de 2005.** 2005. Disponível em: <http://www.planalto.gov.br/ccivil{_}03/{_}ato2004-2006/2005/decreto/d56>. Citado na página 26.

BRENTARI, D. **A prosodic model of sign language phonology**. Cambridge: MIT Press, 1998. 396 p. ISBN 9780262024457. Citado na página 38.

BRITO, L. F. a.; LANGEVIN, R. The sublexical structure of a sign language. **Mathématiques et sciences humaines**, v. 32, n. 125, p. 17–40, 1994. Citado na página 16.

CAPOVILLA, F. C.; RAPHAEL, W. D. **Dicionário enciclopédico ilustrado trilíngüe da língua de sinais brasileira: Sinais de A a L**. 2. ed. São Paulo, SP: EdUSP, 2001. v. 1. 1620 p. ISBN 8531406684. Citado 4 vezes nas páginas 3, 27, 59 e 60.

CAPOVILLA, F. C.; RAPHAEL, W. D. **Dicionário enciclopédico ilustrado trilíngüe da língua de sinais brasileira: Sinais de M a Z**. 2. ed. São Paulo, SP: EdUSP, 2001. v. 2. 1620 p. ISBN 9788531406690. Citado 3 vezes nas páginas 3, 59 e 60.

CARMONA, J. C. C. Dissertação (Metrado em Ensino de Ciências Humanas, Sociais e da Natureza), **A dicionarização de termos em língua brasileira de sinais (Libras) para o ensino de biologia: uma atitude empreendedora**. Londrina: [s.n.], 2015. 172 p. Citado na página 28.

CAVENDER, A.; LADNER, R. E. Ntid international symposium on technology and deaf education. **ACM SIGACCESS Accessibility and Computing**, n. 97, p. 3–13, 2010. ISSN 15582337. Disponível em: <http://portal.acm.org/citation.cfm?doid=1873532.1873533>. Citado 4 vezes nas páginas 29, 38, 47 e 48.

CAVENDER, A. C.; BIGHAM, J. P.; LADNER, R. E. Classinfocus: Enabling improved visual attention strategies for deaf and hard of hearing students. **Proceeding of the eleventh international ACM SIGACCESS conference on Computers and accessibility - ASSETS '09**, p. 67, 2009. Disponível em: <http://dl.acm.org/citation.cfm?id=1639642.1639656>. Citado na página 29.

CAVENDER, A. C. et al. Asl-stem forum: enabling sign language to grow through online collaboration. In: **28th international conference on Human factors in computing systems - CHI '10**. Atlanta, Georgia, USA: ACM, 2010. (CHI'10), p. 2075. ISBN 9781605589299. Disponível em: <http://dl.acm.org.offcampus.lib.washington.edu/citation.cfm?id=1753326.1753642>. Citado na página 38.

CHAPELLE, O. et al. Large scale validation and analysis of interleaved search evaluation. **ACM Transactions on Information Systems (TOIS)**, ACM Press, v. 30, n. 1, p. 6:1–6:41, 2012. ISSN 10450823. Disponível em: <http://doi.acm.org/10.1145/2094072.2094078>. Citado na página 59.

CHUKLIN, A. et al. A comparative analysis of interleaving methods for aggregated search. **ACM Transactions on Information Systems**, v. 33, n. 2, p. 147–156, 2015. Citado na página 59.

CORMIER, K. et al. From corpus to lexical database to online dictionary : Issues in annotation of the bsl corpus and the development of bsl signbank bsl lexical database (bld) from bld to bsl signbank. In: **5th Workshop on the representation and processing of sign languages: Interactions between corpus and lexicon: part of the 8th International Conference on Language Resources and Evaluation**. Turkey, Istanbul: [s.n.], 2012. p. 7–12. Citado 3 vezes nas páginas 37, 38 e 48.

COSTA, A. C. d. R. **The SignNet Project: IT Tools for the Written Deaf Culture**. Pelotas - RS, 2001. 7 p. Disponível em: <http://www.signwriting.org/archive/docs3/sw0215-BR-SignNet-Proj-2001.pdf>. Citado na página 59.

COSTA, A. C. d. R.; DIMURO, G. P. Signwriting and swml: Paving the way to sign language processing. In: STREITER (Ed.). **Traitement Automatique des Langues de Signes, Workshop on Minority Languages**. Batz-sur-Mer: [s.n.], 2003. (TALN), p. 11–14. Citado na página 23.

COSTA, A. C. d. R. et al. Um convite ao processamento de línguas de sinais. In: **TIL 2004 - Workshop em Tecnologia da Informação e da Linguagem Humana**. Salvador, BA: SBC, 2004. p. 1–10. Citado na página 23.

COSTA, V. H. S. da. Dissertação (Mestrado em Linguística), **Iconicidade e produtividade na língua brasileira de sinais**. Florianópolis, SC: [s.n.], 2012. 96 p. Citado na página 42.

COSTAGLIOLA, G. et al. Improving shape context matching for the recognition of sketched symbols. In: **DMS'11 - The 17th International Conference on Distributed Multimedia Systems**. [S.l.: s.n.], 2011. p. 289–294. ISBN 1891706306. Citado na página 52.

CURRIE, A.-M. P. G.; MEIER, R. P.; WALTERS, K. A crosslinguistic examination of the lexicons of four signed languages. In: MEIER, R. P.; CORMIER, K.; QUINTO-POZOS, D. (Ed.). **Modality and structure in signed and spoken languages**. Cambridge: Cambridge University Press, 2002. p. 224–236. Disponível em: <http://ebooks.cambridge.org/ref/id/CBO9780511486777A021>. Citado na página 26.

DEUCHAR, M. **British Sign Language**. 1. ed. New York, NY: Routledge, 2013. 246 p. ISBN 978-0415867948. Citado na página 20.

DIAS, M. A. L.; MALHEIROS, M. d. G. Estudo de técnicas de radicalização para a língua portuguesa. In: **1º WORKSHOP DE ENGENHARIA E TECNOLOGIA /1ª Semana Acadêmica das Engenharias/III Simpósio de Computação, Informática e Tecnologia do Vale do Taquari**. Lajeado, RS: Univates, 2005. p. 1–14. Disponível em: <http://ensino.univates.br/{~}abadia/artigo-wet.p>. Citado na página 33.

DIAS, T.; PEDROSO, C.; ROCHA, J. Uma análise sobre o ensino de libras para familiares ouvintes de surdos. In: LIMA, R. d. C. P.; CARVALHO, G. M. F. (Ed.). **Sujeito, escola e representações**. [S.l.]: Insular, 2006. p. 1–16. ISBN 8574743305. Citado na página 3.

DINIZ, H. G. Dissertação (Mestrado em Linguística), **A História da Língua de Sinais Brasileira (Libras)**. Florianópolis, SC: [s.n.], 2010. 112 p. Citado na página 30.

DRIGAS, A. S. et al. E-learning environment for deaf people in the e-commerce and new technologies sector. **WSEAS Transactions on Information Science and Applications**, v. 1, n. 5, p. 1189–96, 2004. ISSN 1790-0832. Citado na página 29.

EMMOREY, K. **Language, cognition, and the brain: Insights from sign language research**. Nahwash, NJ: Psychology Press, 2001. 408 p. ISBN 978-0805833997. Citado 2 vezes nas páginas 12 e 13.

EMMOREY, K. et al. How sensory-motor systems impact the neural organization for language: Direct contrasts between spoken and signed language. **Frontiers in Psychology**, v. 5, n. may, p. 1–13, 2014. ISSN 16641078. Citado na página 14.

ESTADO DE MINAS GERAIS. **Lei 10.379, de 10/01/1991**. 1991. Citado na página 26.

FAULSTICH, E. **Nota lexical Sinal-Termo**. Brasília, DF, 2014. Disponível em: <www.centrolexterm.com.br>. Citado na página 11.

FELIPE, T. A. Os processos de formação de palavra na libras. **ETD - Educação Temática Digital**, v. 7, n. 2, p. 200–217, 2006. Citado 6 vezes nas páginas 8, 17, 18, 40, 41 e 49.

FELIPE, T. A. **Libras em Contexto : Curso Básico : Livro do Estudante**. 8. ed. Rio de Janeiro, RJ: WalPrint Gráfica e Editora, 2007. 168 p. ISBN 8599091018. Citado 2 vezes nas páginas 12 e 30.

FELIPE, T. A. The verbal-visual discourse in brazilian sign language – libras. **Bakhtiniana. Revista de Estudos do Discurso**, v. 8, n. 2, p. 67–89, 2013. ISSN 2176-4573. Citado na página 14.

FERNANDES, J. D. C. Introdução à semiótica. In: ALDRIGUE, A. C. d. S.; LEITE, J. E. R. (Ed.). **LIinguagens- Usos e reflexões**. João Pessoa, PB: [s.n.], 2011. cap. Introdução, p. 161–185. ISBN 972-762-088-4. Citado na página 19.

FERREIRA, S. P. A.; DIAS, M. d. G. B. B. the school and the reading learning process. **Psicologia em Estudo**, v. 7, n. 1, p. 39–49, 2002. Citado na página 2.

FRYDRYCH, L. A. K. Rediscutindo as noções de arbitrariedade e iconicidade: implicações para o estatulo linguístico das línguas de sianis. **ReVEL**, v. 10, n. 19, p. 281–294, 2012. Citado 2 vezes nas páginas 44 e 45.

GARDIES, R. **Compreender o cinema e as imagens**. Lisboa: TEXTO & GRAFIA, 2008. 52 p. ISBN 978-989-95689-8-3. Citado na página 33.

GLICKMAN, N. Do you hear voices? problems in assessment of mental status in deaf persons with severe language deprivation. **Journal of Deaf Studies and Deaf Education**, v. 12, n. 2, p. 127–147, 2007. ISSN 10814159. Citado na página 8.

GOMES, V.; TEIXEIRA. A iconicidade e arbitrariedade na libras. **Revista Philologus**, v. 20, n. 58, p. 91–98, 2014. Citado 4 vezes nas páginas 30, 44, 45 e 49.

GUIMARÃES, C. **Arquitetura pedagógica computacional para interações intelectuais entre crianças surdas e pais não-surdos em Libras e português**. 161 p. Tese ((Doutorado em Informática)) — Universidade Federal do Paraná, Curitiba, PR, 2013. Citado 3 vezes nas páginas 5, 42 e 43.

HANKE, T.; KONRAD, R.; SCHWARZ, A. Glosslexer: A multimedia lexical database for sign language dictionary compilation. **Sign Language & Linguistics**, v. 4, n. 1-1, p. 171–189, 2001. Citado 2 vezes nas páginas 37 e 48.

HANKE, T.; STORZ, J. ilex – a database tool for integrating sign language corpus linguistics and sign language lexicography. In: **Sign Language and Linguistics**. Marrakech, Morocco: European Language Resources Association (ELRA), 2008. (Irec), p. 64–67. Citado 2 vezes nas páginas 37 e 48.

HIBBARD, E. S.; FELS, D. I. The vlogging phenomena: a deaf perspective. **Assets (Xiii)**, p. 59, 2011. Disponível em: <http://dl.acm.org/citation.cfm?doid=2049536.2049549>. Citado na página 16.

HOCKETT, C. F.; HOCKETT, C. D. The origin of speech. **Scientific American**, Scientific American, a division of Nature America, Inc., v. 203, n. 3, p. 88–97, 1960. ISSN 00368733, 19467087. Disponível em: <http://www.jstor.org/stable/24940617>. Citado na página 12.

HOLGATE, S. A. Strength in disability. **Science**, AAAS, v. 347, n. 6229, p. 1510–1510, 2015. Citado 2 vezes nas páginas 38 e 39.

HOUAISS, A.; Mauro de Salles Villar. **Dicionário Houaiss Da Língua Portuguesa - Novo**. 1. ed. Rio de Janeiro, RJ: Objetiva, 2009. 2048 p. ISBN 978-8573029635. Citado na página 27.

HUDDLESTON, R.; PULLUM, G. K. The cambridge grammar of the english language. p. 1–23, 2002. ISSN 0013-838X. Citado 2 vezes nas páginas 12 e 13.

IBGE, I. B. d. G. e. E. **Censo Demográfico - Características gerais da população, religião e pessoas com deficiência**. Rio de Janeiro, RJ, 2010. 1–211 p. Disponível em: <http://biblioteca.ibge.gov.br/visualizacao/periodicos/94/cd{_}2010{_}religiao{_}defici>. Citado na página 49.

ITU-T. **Recommendation P.910 - Subjective video quality assessment methods for multimedia applications**. [S.l.], 2008. 1–42 p. Disponível em: <papers2://publication/uuid/F9083E1E-CFD2-4069-88C1-4AC8B8A224DA>. Citado na página 60.

JIANG, H.; DREW, M. S.; LI, Z.-n. Action detection in cluttered video with successive convex matching. **Circuits and Systems for Video Technology, IEEE Transactions on**, v. 20, n. 1, p. 50–64, 2010. Citado na página 52.

JOHNSTON, T. **Auslan: The Sign Language of the Australian Deaf Community**. 267 p. Tese ((Doutorado em Linguística da Filosofia)) — University of Sydney, Sydney, Austrália, 1989. Citado na página 11.

JOHNSTON, T. A. Language standardization and signed language dictionaries. **Sign Language Studies**, Gallaudet University Press, v. 3, n. 4, p. 431–468, 2003. Citado 2 vezes nas páginas 37 e 48.

KANIS, J. et al. A methodology for automatic sign language dictionary creation. In: **Universal Learning Design**. Brno, Czech Republic: [s.n.], 2011. p. 1–9. Disponível em: <http://www.kky.zcu.cz/en/publications/KanisJakub{_}2011{_}AMethodol>. Citado na página 38.

KELLY, R. R.; GAUSTAD, M. G. Deaf college students' mathematical skills relative to morphological knowledge, reading level, and language proficiency. **Journal of Deaf Studies and Deaf Education**, Oxford Univ Press, v. 12, n. 1, p. 25–37, 2006. ISSN 10814159. Citado 3 vezes nas páginas 39, 47 e 48.

KOHONEN, T. Median strings. **Pattern Recognition Letters**, Elsevier Science, v. 3, n. 5, p. 309–313, 1985. ISSN 01678655. Citado na página 53.

KOLLER, O.; NEY, H.; BOWDEN, R. May the force be with you: Force-aligned signwriting for automatic subunit annotation of corpora. In: **Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on**. Shanghai, China: IEEE, 2013. p. 1–6. ISBN 9781467355452. Citado na página 23.

KONSTAN, J. a. et al. Grouplens: applying collaborative filtering to usenet news. **Communications of the ACM**, Communications of the ACM, v. 40, n. 3, p. 77–87, 1997. ISSN 00010782. Citado na página 39.

KRIEGER, M. da G.; FINATTO, M. J. B. **Introdução à Terminologia: teoria e prática**. São Paulo, SP: Editora Contexto, 2004. 224 p. ISBN 978-8572442589. Citado na página 27.

KUSHALNAGAR, R. S. A comparison of p2p approaches in handling multiple video perspective streams by mobile devices to support a more inclusive learning environment. In: **2010 8th IEEE International Conference on Pervasive Computing and Communications Workshops, PERCOM Workshops 2010**. Mannheim, Germany: IEEE, 2010. p. 855–856. ISBN 9781424466054. Citado na página 29.

KUSHALNAGAR, R. S.; CAVENDER, A. C.; PÂRIS, J.-f. Multiple view perspectives: Improving inclusiveness and video compression in mainstream classroom recordings. In: **Proceedings of the 12th international ACM SIGACCESS conference on Computers and accessibility ASSETS 10**. Orlando, Florida, USA: ACM, 2010. p. 123–130. ISBN 9781605588810. Disponível em: <http://portal.acm.org/citation.cfm?doid=1878803.1878827>. Citado na página 29.

LADNER, R. E.; CAVENDER, A. C. **ASL-STEM**. 2009. University of Washington. Disponível em: <aslstem.cs.washington.edu>. Acesso em: 19/09/2015. Citado na página 38.

LAMBERG, W. J.; LAMB, C. E. **Reading instruction in the content areas**. [S.l.]: Houghton Mifflin Harcourt (HMH), 1980. v. 1. 412 p. ISBN 9780395306482. Citado na página 36.

LANG, H. Higher education for deaf students: Research priorities in the new millennium. **Journal of Deaf Studies and Deaf Education**, Oxford University Press, v. 7, n. 4, p. 267–80, 2002. ISSN 1465-7325. Disponível em: <www.ncbi.nlm.nih.gov/pubmed/15451865>. Citado na página 3.

LANG, H. G.; EGELSTON-DODD, J.; SACHS, M. C. Science education for hearing-impaired students in the eighties: Priorities and projections. **American Annals of the Deaf**, Gallaudet University Press, v. 128, n. 6, p. 801–808, 1983. ISSN 0002726X. Citado na página 3.

LANG, H. G. et al. A study of technical signs in science: Implications for lexical database development. **Journal of Deaf Studies and Deaf Education**, Oxford University Press, v. 12, n. 1, p. 65–79, 2007. ISSN 10814159. Citado 4 vezes nas páginas 2, 3, 35 e 36.

LEVENSHTEIN, V. I. Binary codes capable of correcting deletions, insertions, and reversals. **Soviet Physics Doklady**, v. 10, n. 8, p. 707–710, 1966. ISSN 00385689. Citado na página 53.

LIMA, V. L. d. S. e. **Língua de Sinais: proposta terminológica para a área de desenho arquitetônico**. 272 p. Tese ((Doutorado em Linguística Teórica e Descritiva)) — Universidade Federal de Minas Gerais, Belo Horizonte, MG, 2014. Citado 16 vezes nas páginas x, 1, 3, 9, 13, 14, 17, 18, 20, 26, 27, 40, 46, 59, 61 e 66.

LIRA, G. d. A.; SOUZA, T. A. F. d. **Dicionário da Língua Brasileira de Sinais V3**. 2011. Acessibilidade Brasil. Disponível em: <http://www.acessibilidadebrasil.org.br/libras_3>. Acesso em: 01/08/2017. Citado na página 27.

LU, P.; HUENERFAUTH, M. Data-driven synthesis of spatially inflected verbs for american sign language animation. **ACM Transactions on Accessible Computing**, v. 4, n. 1, p. 1–29, 2011. ISSN 19367228. Disponível em: <http://www.scopus.com/inward/record.url?eid=2-s2.0-81755176764{&}partnerID=tZOtx>. Citado na página 28.

LYRAS, D. P.; SGARBAS, K. N.; FAKOTAKIS, N. D. Using the levenshtein edit distance for automatic lemmatization: A case study for modern greek and english. In: **Proceedings - International Conference on Tools with Artificial Intelligence**. Patras, Greece: IEEE Computer Society Press, 2007. (ICTAI, v. 2), p. 428–435. ISBN 076953015X. ISSN 10823409. Citado na página 53.

MAIA, O. B. **Um sistema de gerenciamento da qualidade de experiência orientada à transmissão de vídeos para dispositivos móveis em redes sem fio**. 175 p. Tese ((Doutorado em Engenharia Elétrica)) — Universidade Federal de Minas Gerais, Belo Horizonte, MG, feb 2015. Citado na página 63.

MANNING, C. D.; SCHÜTZE, H. **Foundations of Statistical Natural Language Processing**. Cambridge: MIT Press, 1999. 620 p. ISBN 978-0262133609. Citado na página 32.

MARQUES, P. A. **Dicionário Libras**. 2002. Super Modular. Disponível em: <www.dicionariolibras.com.br/>. Acesso em: 2015-09-15. Citado na página 3.

MARSCHARK, M.; HAUSER, P. C. Cognitive underpinnings of learning by deaf and hard-of-hearing students. In: **Deaf Cognition: Foundations and outcomes**. New York, NY, USA: Oxford University Press, 2008. v. 1973, cap. Cognitive, p. 3–23. Citado na página 2.

NAKAMURA, K. **The Deaf Resource Library**. 1995. Disponível em: <http://www.deaflibrary.org/asl.html>. Acesso em: 10/01/2018. Citado 2 vezes nas páginas 12 e 13.

NASCIMENTO, S. P. d. F. d. **Representações lexicais da língua de sinais brasileira : uma proposta lexicográfica**. 290 p. Tese ((Doutorado em Linguística)) — Universidade de Brasília, Brasília, 2009. Citado 9 vezes nas páginas 4, 8, 17, 18, 19, 30, 40, 41 e 45.

NUNES, V. F. Iconicidade e corporificação em sinais de libras : uma abordagem cognitiva. In: CARVALHO, D. R. Gisele de; VASCONCELLOS, Z. (Ed.). **Linguagem: Teoria, Análise e Aplicações (7)**. Rio de Janeiro, RJ: Programa de Pós-Graduação em Letras da UERJ, 2009. p. 245 – 253. ISBN 978-85-64315-06-8. Disponível em: <http://www.pgletras.uerj.br/linguistica/linguagem07.html>. Citado 3 vezes nas páginas 20, 42 e 44.

ORENGO, V.; HUYCK, C. A stemming algorithmm for the portuguese language. In: **Proceedings Eighth Symposium on String Processing and Information Retrieval**. Laguna de San Rafael: IEEE Computer Society Press, 2001. (SPIRE), p. 186–193. ISBN 0-7695-1192-9. Citado 4 vezes nas páginas viii, 33, 45 e 53.

PARTON, B. S. Sign language recognition and translation: A multidisciplined approach from the field of artificial intelligence. **Journal of Deaf Studies and Deaf Education**, v. 11, n. 1, p. 94–101, 2006. ISSN 10814159. Citado na página 12.

PERLMAN, M.; DALE, R.; LUPYAN, G. Iconicity can ground the creation of vocal symbols. **Royal Society Open Science**, v. 2, n. 8, p. 150152, 2015. ISSN 2054-5703. Disponível em: <http://rsos.royalsocietypublishing.org/content/2/8/150152.abstract>. Citado 3 vezes nas páginas 43, 44 e 49.

PERNISS, P.; VIGLIOCCO, G. The bridge of iconicity: from a world of experience to the experience of language. **Philosophical Transactions of the Royal Society B: Biological Sciences**, v. 369, n. 1654, p. 20140179–20140179, 2014. ISSN 0962-8436. Disponível em: <http://rstb.royalsocietypublishing.org/cgi/doi/10.1098/rstb.2014.0179>. Citado 2 vezes nas páginas 16 e 43.

PIVETTA, E. M.; SAITO, D. S.; FLOR, S. Análise semiótica da língua de sinais. **ACTA SEMIOTICA ET LINGVISTICA**, v. 18, n. 2, p. 260–271, 2013. Citado 2 vezes nas páginas 14 e 44.

PIZZIO, A. L. **A tipologia linguística e a língua de sinais brasileira: elementos que distinguem nomes de verbos**. 237 p. Tese ((Doutorado em Linguística)) — Universidade Federal de Santa Catarina, Florianópolis, SC, 2011. Disponível em: <https://repositorio.ufsc.br/xmlui/bitstream/handle/123456789/95954/291547.pdf?sequence=1>. Citado na página 3.

PONTES, A. L. Terminologia científica: o que é e como se faz. **Revista de Letras**, v. 1, n. 19, p. 44–51, 1997. Citado na página 27.

QUADROS, R. M. de; KARNOPP, L. B. **Língua de sinais brasileira - Estudos linguísticos**. São Paulo, SP: Artmed, 2004. 222 p. ISBN 85-363-0308-5. Citado 7 vezes nas páginas 13, 15, 16, 17, 18, 40 e 49.

QUADROS, R. M. de; PERLIN, G. **Estudos Surdos II**. 1. ed. Petrópolis, RJ: Arara Azul, 2007. 268 p. ISBN 978-85-89002-21-9. Citado na página 2.

REDDEN, M. R.; DAVIS, C. A.; BROWN, J. W. Science for handicapped students in higher education: Barriers, solutions and recommendations. **American Association for the Advancement of Science**, ERIC, n. ED188381, p. 94, 1978. Citado na página 3.

RIZZOLATTI, G.; ARBIB, M. A. Language within our grasp. **Trends in Neurosciences**, v. 21, n. 5, p. 188–194, 1998. ISSN 01662236. Citado na página 12.

ROBERSON, L. Integration of computers and related technologies into deaf education teacher preparation programs. **American annals of the deaf**, v. 146, n. 1, p. 60–66, 2001. ISSN 0002-726X. Citado 2 vezes nas páginas 29 e 35.

ROCHE, E.; SCHABES, Y. **Finite-state language processing**. Cambridge: MIT Press, 1997. 482 p. ISBN 978-0262527705. Citado na página 32.

RODRIGUES, A.; MUNHOZ, A. V.; HATTGE, M. D. Corpo surdo e suas singularidades. **Reflexão e Ação**, v. 23, n. 3, p. 104, 2015. ISSN 1982-9949. Disponível em: <http://online.unisc.br/seer/index.php/reflex/article/view/6083>. Citado na página 2.

ROHALY, A. M. et al. Video quality experts group: Current results and future directions. In: **Visual Communications and Image Processing 2000**. International Society for Optics and Photonics, 2000. p. 742–753. ISBN 6139987822. ISSN 0277786X. Disponível em: <http://spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.386632>. Citado na página 60.

SAUSSURE, F. de. **Curso de Lingüística Geral**. 27. ed. São Paulo, SP: Cultrix, 1970. 312 p. ISBN 978-8531601026. Citado na página 19.

SAUSSURE, F. de. **Curso de Lingüística Geral**. São Paulo, SP: Editora Cultrix, 2006. 278 p. ISBN 978-85-316-0102-6. Citado na página 44.

SCHERMER, G. M. From variant to standard: An overview of the standardization process of the lexicon of sign language of the netherlands over two decades. **Sign Language Studies**, Gallaudet University Press, v. 3, n. 4, p. 469–486, 2003. ISSN 1533-6263. Citado 2 vezes nas páginas 37 e 48.

SILVA, F. I. da. Dissertação (Mestrado em Educação), **Analisando o processo de leitura de uma possível escrita da língua brasileira de sinais : Signwriting**. Florianópolis, SC: [s.n.], 2009. 116 p. Citado na página 44.

SIMÕES, D. **Iconicidade Verbal : Teoria e Prática**. Rio de Janeiro, RJ: Dialogarts, 2009. ISBN 9788586837623. Citado 3 vezes nas páginas 20, 30 e 42.

SLEVINSKI, S. E. **SignPuddle Online v2.0**. 2012. Pittsburgh United for Deaf Literacy. Disponível em: <www.signbank.org/signpuddle2.0/index.php?ui=12{&}sgn>. Acesso em: 20/09/2015. Citado na página 3.

SLEVINSKI, S. E. **Formal SignWriting**. 2013. Center for Sutton Movement Writing. Disponível em: <slevinski.github.io/SuttonSignWriting/assets/draft-slevinski-formal-signwriting.html>. Acesso em: 02/01/2015. Citado 4 vezes nas páginas 23, 24, 25 e 26.

SOFIATO, C. G.; REILY, L. Justaposições: o primeiro dicionário brasileiro de língua de sinais e a obra francesa que serviu de matriz. **Revista Brasileira de Educação Especial**, v. 18, n. 4, p. 569–586, 2012. Citado na página 26.

SOUSA, S. F. D.; SILVEIRA, H. E. Terminologias químicas em libras: A utilização de sinais na aprendizagem de alunos surdos. **Revista Química Nova na Escola**, v. 33, n. 1, p. 37–48, 2010. Citado 2 vezes nas páginas 9 e 28.

STADLER, J. P.; FILIETAZ, M. R. P.; HUSSEIN, F. R. G. S. O ensino bilíngue libras-português na disciplina de química : a importância do uso de sinais específicos resultados e discussão. In: **XVI ENEQ/X EDUQUI - EI - Ensino de Química e inclusão**. Salvador, BA: UFBA, 2012. ISSN 2179-5355. Citado 3 vezes nas páginas 3, 8 e 9.

STADLER, J. P.; FILIETAZ, M. R. P.; Hussein, Fabiana, R. G. e. S. Investigação de terminologias científicas de química em língua ‘ brasileira de sinais em escola bilíngue de curitiba, brasil. **Revista Tecné, Episteme y Didaxis: TED**, p. 234–241, 2014. ISSN 2323-0126. Citado 2 vezes nas páginas 8 e 28.

STEELS, L. Modeling the cultural evolution of language. **Physics of Life Reviews**, Elsevier B.V., v. 8, n. 4, p. 339–356, 2011. ISSN 15710645. Disponível em: <http://dx.doi.org/10.1016/j.plrev.2011.10.014>. Citado na página 49.

STIEHL, D. et al. Towards a signwriting recognition system. **Proceedings of the International Conference on Document Analysis and Recognition, ICDAR**, v. 2015-November, p. 26–30, 2015. ISSN 15205363. Citado na página 57.

STOKOE, W. C. Sign language structure: an outline of the visual communication systems of the american deaf. **Journal of deaf studies and deaf education**, Oxford University Press, v. 10, n. 1, p. 3–37, jan 2005. ISSN 1081-4159. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/15585746>. Citado na página 15.

STROBEL, K. **As imagens do outro sobre a cultura surda**. 2. ed. Florianópolis, SC: Editora UFSC, 2008. 118 p. ISBN 9788532804280. Citado na página 4.

STUMPF, M. R. **Aprendizagem de Escrita de Língua de Sinais pelo sistema SignWriting: Línguas de Sinais no papel e no computador**. 329 p. Tese ((Doutorado em Informática na Educação)) — Universidade Federal do Rio Grande do Sul, Porto Alegre, 2005. Citado 8 vezes nas páginas 2, 4, 8, 11, 21, 35, 40 e 43.

STUMPF, M. R.; OLIVEIRA, J. S. de; MIRANDA, R. D. **Glossário Libras**. 2006. Universidade Federal de Santa Catarina. Disponível em: <www.glossario.libras.ufsc.br/>. Acesso em: 15/09/2015. Citado na página 3.

SUTTON, V. **SignWriting Site**. 1994. Valerie Sutton. Disponível em: <http://www.signwriting.org/>. Acesso em: 19/01/2015. Citado na página 22.

SUTTON, V. Writing sign language on your computer. **ComputorEdge**, San Diego, CA, v. 24, n. 07, p. 16–18, jul 2006. Citado na página 23.

SUTTON, V. J. **Lessons in SignWriting Textbook**. 4. ed. La Jolla, CA: The SignWriting Press, 2014. v. 1. 220 p. ISBN 091433655X. Citado 2 vezes nas páginas 56 e 57.

TAKAHIRA, A. G. R. Questões sobre compostos e morfologia da libras. **Revista Estudos Linguísticos**, v. 41, n. 1, p. 262–276, 2012. ISSN 1413-0939. Citado 2 vezes nas páginas 40 e 41.

TEMOTEO, J. G. **Lexicografia da Língua de Sinais Brasileira do Nordeste**. 252 p. Tese ((Doutorado em Psicologia Experimental)) — Universidade de São Paulo, São Paulo, SP, 2012. Citado na página 42.

THIESSEN, S. M. Dissertação (Master of Arts), **A Grammar of SignWriting**. 2011. 300 p. Citado na página 57.

TOLAR, T. D. et al. The development of the ability to recognize the meaning of iconic signs. **Journal of Deaf Studies and Deaf Education**, v. 13, n. 2, p. 225–240, 2008. ISSN 10814159. Citado 2 vezes nas páginas 14 e 45.

UNION, T. S. S. o. I. T. **Method for objective measurments of perceived audio quality**. [S.l.], 1998. 1–5 p. Citado 2 vezes nas páginas 59 e 60.

VESEL, J.; ROBILLARD, T. Increasing access to technical science vocabulary through use of universally designed signing dictionaries. **Universal Design in Higher Education: Promising Practices**, University of Washington, Washington, DC, p. 1–6, 2013. Citado 2 vezes nas páginas 39 e 48.

VESEL, J.; ROBILLARD, T. Teaching mathematics vocabulary with an interactive signing math dictionary. **Journal of Research on Technology in Education**, Taylor & Francis, v. 45, n. 4, p. 361–389, 2013. ISSN 15391523. Citado 3 vezes nas páginas 39, 47 e 48.

VOGLER, C.; GOLDENSTEIN, S. Toward computational understanding of sign language. **Technology and Disability**, IOS PRess, v. 20, n. 2, p. 2008, 2008. ISSN 1055-4181. Citado na página 38.

WANG, Y. Inquiry-based science instruction and performance literacy for students who are deaf or hard of hearing. **American Annals of the Deaf**, Gallaudet University Press, v. 156, n. 3, p. 239–254, 2011. ISSN 0002-726X. Citado na página 35.

WELLENS, P. et al. Fluid construction grammar for historical and evolutionary linguistics. **Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics**, p. 127–132, 2013. Citado na página 49.

WESCOTT, R. W. Linguistic iconism. **Linguistic Society of America**, v. 47, n. 2, p. 416–428, 1971. Disponível em: <http://www.jstor.org/stable/412089>. Citado na página 20.

WEST, D.; SUTTON-SPENC, R. Shared thinking processes with four deaf poets: A window on "the creative"in "creative sign language". **Sign Language Studies**, v. 12, n. 2, p. 188–211, 2012. ISSN 03021475. Disponível em: <http://madonnaezp.liblime.com/login?url=http://search.proquest.com/docview/1010283606?accountid=27927>. Citado na página 29.

ZWITSERLOOD, I. Sign language lexicography in the early 21st century and a recently published dictionary of sign language of the netherlands. **International Journal of Lexicography**, v. 23, n. 4, p. 443–476, 2010. ISSN 09503846. Citado na página 27.

Apêndices

APÊNDICE A – Parecer do CEP

PARECER CONSUBSTANCIADO DO CEP

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: Sistema Computacional de Apoio à Criação e Preservação Terminológica em Línguas de Sinais: Neologismos em Ciência, Tecnologia, Engenharia e Matemática

Pesquisador: Celso Luiz de Souza
Área Temática:
Versão: 2
CAAE: 53435716.0.0000.5588
Instituição Proponente: Instituto Federal de Educação Ciência e Tecnologia do Sudeste de Minas
Patrocinador Principal: Financiamento Próprio

DADOS DO PARECER

Número do Parecer: 1.635.736

Apresentação do Projeto:
O projeto possui boa apresentação e apresenta a seguinte hipótese:"A Libras - Língua Brasileira de Sinais consolidou-se no Brasil através de um plano linguístico que surgiu a partir da Lei Federal 10.436, de abril de 2002 e do subsequente Decreto Federal 5.695, de dezembro de 2005. Nesse novo cenário, vários esforços têm sido empregados, não só pelo
governo brasileiro, mas também por várias associações surdas para implementar linguística prática e políticas que possam garantir uma efetiva
inclusão social dos surdos no sistema educacional brasileiro (Quadros 2012). O trabalho atual é parte desta mobilização social nacional importante.Há uma carência por obras terminográficas bilíngues (Língua Portuguesa e Língua Brasileira de Sinais) e isso é percebido em várias áreas, na
acadêmica, na tecnológica e na cultural, e em vários níveis do aprendizado. (Lima 2014) apresenta uma preocupação usual entre os estudantes
surdos, intérpretes de Libras e educadores que é a falta de sinais para os termos científicos que são utilizados ao longo do ensino das diversas disciplinas.Os termos gerados pelo sistema, pode ser avaliados de forma similar ao desenvolvidos por grupo de usuários da língua." E no que diz respeito à metodologia o pesquisador a descreve da seguinte forma: "A metodologia principal é

empregada para o desenvolvimento de um gerador de novos sinais para línguas de sinais, em específico, como estudo de caso, a Libras. A ideia é que esse gerador de sinais possa apoiar a produção terminológica das áreas de ciência, tecnologia, engenharia e matemática. Nesse âmbito, será realizado um estudo de caso específico na disciplina de Física. No entanto, essa carência terminológica está presente em todas as áreas."

Objetivo da Pesquisa:

O objetivo da pesquisa "consiste em desenvolver um sistema computacional que tem como propósito subsidiar a criação e a preservação de sinais para termos técnicos em ciência, tecnologia, engenharia e matemática (CTEM) em línguas de sinais."

Avaliação dos Riscos e Benefícios:

Apresenta os riscos e benefícios da seguinte maneira: "Riscos:O sistema gerar sinais que não são aplicáveis a Libras.Benefícios:Desenvolvimento de dicionários terminológicos para língua de sinais."

Comentários e Considerações sobre a Pesquisa:

Trata-se de pesquisa muito relevante no que diz respeito ao desenvolvimento de sinais que auxiliarão na inserção de surdos-mudos no mercado de trabalho nas áreas de ciência, tecnologia,engenharia e matemática.

Considerações sobre os Termos de apresentação obrigatória:

Os termos de apresentação obrigatória foram corretamente apresentados.

Conclusões ou Pendências e Lista de Inadequações:

Não há pendências.

Considerações Finais a critério do CEP:

Diante do exposto, o Comitê de Ética em Pesquisa com Seres Humanos CEP/IF SUDESTE MG, de acordo com as atribuições definidas na Res. CNS 466/12, após análise, manifesta-se FAVORÁVEL à aprovação do projeto de pesquisa referenciado.

1. Comunicar toda e qualquer alteração do projeto e termo de consentimento livre e esclarecido.
2. Comunicar imediatamente ao Comitê qualquer evento adverso ocorrido durante o desenvolvimento do estudo.
3. Os dados individuais de todas as etapas da pesquisa devem ser mantidos em local seguro por 5 anos para possível auditoria dos órgãos competentes.
4. Conforme definido no artigo 58 do regimento do CEP/IF SUDESTE MG, inciso III, o pesquisador

deve apresentar relatórios semestrais e final da pesquisa.

Este parecer foi elaborado baseado nos documentos abaixo relacionados:

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PB_INFORMAÇÕES_BÁSICAS_DO_PROJETO_646752.pdf	09/06/2016 00:27:33		Aceito
Projeto Detalhado / Brochura Investigador	Projeto.pdf	11/02/2016 00:51:59	Celso Luiz de Souza	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TCLE.doc	08/01/2016 23:56:58	Celso Luiz de Souza	Aceito
Folha de Rosto	FolhaRostoProjeto.pdf	05/01/2016 00:31:04	Celso Luiz de Souza	Aceito

Situação do Parecer:
Aprovado

Necessita Apreciação da CONEP:
Não

JUIZ DE FORA, 14 de Julho de 2016

Assinado por:
Wildson Justiniano Pinto
(Coordenador)