



**UNIVERSIDADE FEDERAL DE ALAGOAS**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM LINGUÍSTICA E LITERATURA**

**SERGIO JOSÉ DA SILVA**

**LINGUÍSTICA DE *CORPUS*: O PROCESSO DE TRANSCRIÇÃO DA VARIEDADE  
DA LIBRAS DA GRANDE MACEIÓ**

**SERGIO JOSÉ DA SILVA**

**LINGUÍSTICA DE *CORPUS*: O PROCESSO DE TRANSCRIÇÃO DA VARIEDADE  
DA LIBRAS DA GRANDE MACEIÓ**

Dissertação apresentada ao Programa de Pós-graduação em Linguística e Literatura, da Faculdade de Letras, da Universidade Federal de Alagoas, como requisito para qualificação de Mestrado.

Orientador: Prof. Dr. Jair Barbosa da Silva.

Linha de pesquisa: Teoria e Análise Linguística.

Maceió - 2022

**Catálogo na fonte**  
**Universidade Federal de Alagoas**  
**Biblioteca Central**  
**Divisão de Tratamento Técnico**

Bibliotecária: Taciana Sousa dos Santos – CRB-4 – 2062

S586l Silva, Sergio José da.  
Linguística de *corpus*: o processo de transcrição da variedade da libras da grande Maceió/ Sergio José da Silva. – 2022.  
98 f. : il. color.

Orientador: Jair Barbosa da Silva.  
Dissertação (Mestrado em Linguística e Literatura) – Universidade Federal de Alagoas. Programa de Pós-Graduação em Linguística e Literatura. Maceió, 2022.

Bibliografia: f. 96-98.

1. Libras. 2. Linguística de corpus. 3. SignBank. I. Título.

CDU: 81'221.24

## AGRADECIMENTOS

Primeiramente, quero agradecer ao Universo pelo dom da vida, por me dar força e aumentar a minha fé para que eu pudesse superar todas as dificuldades.

Agradeço à minha família, principalmente à minha mãe Rosa e ao meu pai Benício, por toda a dedicação e pelo apoio ao longo dos anos; sem esse suporte eu não teria conseguido chegar onde cheguei e serei eternamente grato por isto; aos meus irmãos por terem me incentivado e me dado o auxílio necessário.

Agradeço imensamente à minha noiva, Jéssica Nobre, por ter sido tão especial e por estar ao meu lado me acompanhando com toda dedicação, preocupação e paciência, sempre presente nos momentos difíceis com uma palavra de incentivo, me dando apoio, torcendo por mim e compreendendo o meu esforço. Agradeço também pelo carinho de sua família, por terem me acolhido e me amparado durante esta trajetória; à minha cunhada Kátia Nobre, por ter me ajudado a organizar o sumário e a colocar em prática algumas regras da ABNT.

Ao meu orientador, Prof. Dr. Jair Barbosa, primeiramente por ter aceitado ser meu orientador; em segundo lugar, por todos os ensinamentos que não se resumiram à correções e orientações da dissertação, mas também pelos conselhos e exemplos que vão além da educação, os quais levarei junto comigo para sempre. Obrigado por me acompanhar durante toda a trajetória, desde a graduação, no projeto *corpus* de Libras, onde realizamos coleta de dados para iniciação científica, até hoje; por sempre depositar confiança em mim, prestar-me assistência, pela dedicação, paciência, incentivo para superar minhas dificuldades, enfim, por todo apoio, que foi essencial para a conclusão desta dissertação. Obrigado pela permissão da convivência diária, a você, a minha profunda gratidão.

À profa. Dra. Ronice Quadros serei sempre agradecido pela sua iniciativa do primeiro projeto de *Corpus* de Libras na UFSC, pela defesa da importância da documentação da Libras no Brasil. Também sou muito grato ao prof. Dr. Jair Barbosa ter aceitado a proposta do projeto de *corpus* de Libras em Maceió e por iniciar a realização da coleta de dados, foi enriquecedor para o corpo acadêmico.

À banca examinadora pela leitura cuidadosa e propositiva. Ao prof. Dr. Bruno Carneiro por ter qualificado a minha dissertação, suas sugestões e orientações complementam imensamente toda a minha pesquisa, me ajudaram bastante a aprimorar o trabalho e cresceram as minhas visões sobre o meu título. Agradeço também ao prof. Dr. Adeilson Sedrins por sua disponibilidade ao aceitar o convite e dar continuidade na banca examinadora, sua presença e contribuição são essenciais para mim.

Ao Humberto Meira, por ter aceitado e contribuído com a interpretação da minha qualificação de mestrado, minha eterna gratidão.

Ao Vinicius Leonel, por seu firme trabalho em tradução, pela paciência, pelos momentos de dedicação, os quais me ajudaram nas correções, traduções em português e pelos ensinamentos durante todos esses dias.

Ao Thiago Bruno, por sua empatia, sempre zeloso e cuidadoso ao aprimorar e me instruir sobre a estrutura da língua portuguesa, me ensinando e esclarecendo as muitas interpretações que me fizeram estudar ainda mais e me auxiliando em reflexões profundas.

Ao Marcos Santos, agradeço imensamente pelo seu apoio e pela preocupação, por ter conseguido uma pessoa quando mais precisei para me auxiliar na estruturação da dissertação, sua ajuda foi essencial e somou muito para a conclusão do meu estudo. A você, minha eterna gratidão.

Aos meus diversos colegas Mestres e Doutores, que compartilharam estudos, pesquisas, tirando as dúvidas em atividades, sempre me auxiliando neste momento de pandemia, especialmente me apoiando; obrigado pela atenção, preocupação e pela disponibilidade de todos, em partilhar conhecimentos, sempre acreditando em mim.

À Miriam Royer, pela sua tutoria e por todo esforço em me auxiliar nas matérias necessárias e nos artigos para pesquisas e estudos, aprimorando o meu conhecimento com sua experiência de transcrição em *ELAN*, sempre interagindo nas perguntas e dúvidas que surgiram; obrigado pela sua dedicação ao treinamento, paciência e incentivo.

Aos transcritores do *Corpus* de Libras da UFAL por terem aceitado minha entrevista com as suas ricas experiências sobre transcrição no *ELAN*, foi muito construtivo para minha pesquisa e tiveram uma importância enorme para o desenvolvimento.

Aos meus professores, em geral, por me acompanharem com toda dedicação, ensinamentos e generosidade durante esses anos, por interagiram com paciência e empenho e me incentivaram e me proporcionaram diversas oportunidades.

A todos os intérpretes, pela qualidade dos trabalhos, pela paciência e pelos valorosos conhecimentos compartilhados, sendo incansavelmente compreensivos e gentis.

A todos os amigos e colegas de turma que participaram direta e indiretamente para a concretização desse sonho, e que em todos esses anos de convivência e troca de aprendizados se tornaram fundamentais, meus mais sinceros agradecimentos.

“Quanto mais sinalizamos, mais a Libras se fortalece.”

**Rodrigo Custódio da Silva, 2020**



Fonte :Disponível em:  
<<https://www.youtube.com/watch?v=omdGQ4dL-VU&t=1s>> Acesso  
em: 19 de junho de 2022.

## RESUMO

Esta pesquisa tem por propósito discutir as metodologias e mecanismos de documentação da Libras, além de elaborar estratégias teórico-metodológicas consistentes para o registro dessa Língua, estabelecer e analisar a descrição linguística a partir da documentação da Libras referente à variedade de Maceió. Apresentados os fundamentos da Linguística de *Corpus* referentes à documentação de línguas, especialmente das línguas de sinais, (Quadros e Karnopp (2004), (Crasborn e Sloetjes, 2008), (Johnston 2009), (Hereweghe e Vermeerberge 2012), (Leite e Quadros 2014), (Fenlon, Schembri, Johnston e Cormier 2015) e (Quadros 2016). Metodologicamente, nós focamos nos problemas encontrados no processo de transcrição do projeto *Corpus* de Libras - Maceió. Para isso, entrevistamos os transcritores surdos para entender quais dificuldades eles tiveram durante o processo de transcrição e quais estratégias desenvolveram para lidar com essas dificuldades. Além disso, listamos os sinais os quais os surdos tiveram dificuldade para tentar entender quais problemas poderiam estar por trás destas dificuldades. Nós mapeamos os problemas e os categorizamos em três tipos: a) problemas sociais, ligados ao contexto social que envolve a pesquisa; b) problemas estruturais, que dizem respeito à estrutura do projeto do *Corpus*; e c) metodológicos, ligados às orientações descritas no manual e o uso do *SignBank*. Os problemas metodológicos foram divididos em quatro tipos: problemas de efeito de modalidade, de variação fonológica, de variação lexical e de limitação vocabular do *SignBank*. Como alternativa para solucionar estes problemas, propusemos a garantia da entrada e permanência de surdos na universidade e em projetos de pesquisa, a criação de um grupo de trabalho para os transcritores se apoiarem e validarem os novos sinais e glosas, formações específicas para aprender conceitos linguísticos que facilitem a distinção entre DV(descrição visual), E (gestos) e sinais estáveis, a criação de um código específico para variação fonológica, qual seja: a adição de letras em ordem alfabética para cada nova variante; a criação de um filtro regional no *SignBank* para adicionar variantes lexicais regionais em cada região, e a constante alimentação de novos sinais no *corpus* a partir de constantes reuniões entre os transcritores do projeto e com projetos de outros estados.

**PALAVRAS-CHAVE:** Libras; *Corpus* de Libras; *SignBank*; Transcrição de dados.

## ABSTRACT

This research aims to discuss Libras documentation methodologies and mechanisms, in addition to developing consistent theoretical-methodological strategies for the registration of this language, establishing and analyzing the linguistic description from the Libras documentation referring to the Maceió variety. The fundamentals of Corpus Linguistics regarding the documentation of languages, especially sign languages, are presented (Quadros and Karnopp (2004), (Crasborn and Sloetjes, 2008), (Johnston 2009), (Hereweghe and Vermeerberge 2012), (Leite and Quadros 2014), (Fenlon, Schembri, Johnston and Cormier 2015) and (Quadros 2016) Methodologically, we focused on the problems encountered in the transcription process of the Corpus de Libras - Maceió project. For this, we interviewed deaf transcribers to understand what difficulties they had during the transcription process and what strategies they developed to deal with these difficulties. In addition, we list the signs that the deaf had difficulty in trying to understand what problems could be behind these difficulties. We mapped the problems and categorized them into three types: a) social problems, linked to the social context surrounding the research; b) structural problems, which concern the structure of the Corpus project; and c) methodological, linked to the guidelines described in the manual and the use of SignBank. The methodological problems were divided into four types: modality effect problems, phonological variation, lexical variation and vocabulary limitation of the signal bank. As an alternative to solve these problems, we proposed guaranteeing the entry and permanence of deaf people in the university and in research projects, the creation of a working group for transcribers to support and validate the new signs and glosses, specific training to learn linguistic concepts that facilitate the distinction between DV (visual description), E (gestures) and stable signs, the creation of a specific code for phonological variation, namely: the addition of letters in alphabetical order for each new variant; the creation of a regional filter in SignBank to add regional lexical variants in each region, and the constant feeding of new signs in the corpus from constant meetings between project transcribers and with projects from other Brazilian states.

**KEY WORDS:** Libras; Libras Corpus; SignBank; Data transcription.

## SUMÁRIO

<b>I. CAPÍTULO - Introdução.....</b>	<b>12</b>
1.1 Objetivos da pesquisa.....	13
1.2 Pergunta de pesquisa.....	13
1.3 O problema da pesquisa.....	13
1.4 Estrutura da dissertação.....	15
<b>II. CAPÍTULO – Fundamentação teórica.....</b>	<b>15</b>
2.1 A Linguística de <i>Corpus</i> .....	16
2.1.1 A Sociolinguística.....	17
2.1.2 Construção de <i>corpus</i> e tecnologias.....	20
2.2 Elaboração de <i>corpora</i> em Línguas de Sinais.....	22
2.3 <i>Corpus</i> de Libras.....	28
2.3.1 A Documentação da Libras.....	28
2.3.2 A Documentação da Libras como Política Linguística.....	33
2.3.3 A transcrição dos dados.....	34
2.3.4 A construção do <i>Corpus</i> da Libras da UFAL.....	36
2.3.4.1 Etapas para a elaboração do <i>corpus</i> .....	37
2.3.4.1.1 Os sujeitos da Pesquisa.....	37
2.3.4.2 A Coleta dos Dados.....	39
2.3.4.3 Seleção dos dados.....	41
2.3.4.4 O Tratamento dos Dados.....	44
2.3.4.4.1 A Edição.....	44
2.3.4.4.2 A Anotação/Transcrição dos dados.....	45
<b>III. CAPÍTULO - Metodologia da análise das transcrições.....</b>	<b>53</b>
3.1 Coleta de dados.....	54
3.1.1 Participantes da pesquisa.....	54
3.1.2 Procedimentos e instrumentos da coleta de dados.....	55
3.1.3 Tratamento dos dados (transcrições).....	56
3.1.4 Resultados.....	56
3.1.5 Análise e discussão.....	56
<b>IV. CAPÍTULO - Resultados.....</b>	<b>57</b>
4.1 Relato de experiência sobre a coleta de dados.....	57
4.2 Entrevista com os transcritores.....	58
4.2.1 Resposta ao questionário.....	58
<b>V. CAPÍTULO - Discussão.....</b>	<b>69</b>
5.1 Problemas de transcrição.....	69
5.1.1 Problemas sociais.....	69
5.1.2 Problemas estruturais.....	70

5.1.3 Problemas metodológicos.....	71
5.1.3.1 O efeito da modalidade.....	72
5.1.3.2 Variação fonológica.....	76
5.1.3.3 Variação lexical.....	82
5.1.3.4 Limitação vocabular do <i>SignBank</i> .....	84
5.2 Caminhos para solucionar os problemas de transcrição.....	87
5.2.1 Caminhos para os problemas sociais.....	87
5.2.2 Caminhos para os problemas estruturais.....	87
5.2.3 Caminhos para os problemas metodológicos.....	89
5.2.3.1 Caminhos para os problemas de efeito de modalidade.....	89
5.2.3.2 Caminhos para os problemas de variação fonológica.....	91
5.2.3.3 Caminhos para os problemas de variação lexical.....	93
5.2.3.4 Caminhos para os problemas de limitação do <i>SignBank</i> .....	94
<b>VI. CAPÍTULO - Considerações finais.....</b>	<b>97</b>
<b>REFERÊNCIAS.....</b>	<b>98</b>

## I. CAPÍTULO - Introdução

A Língua Brasileira de Sinais – Libras - há muito é usada pela comunidade surda no Brasil, mesmo antes de ser assim denominada. No entanto, o registro em vídeo dessa língua só passa a ser amplamente disponibilizado a partir do uso das redes sociais, como *Facebook*, *IMO*, *Skype*, *WhatsApp*, *Twitter*, *Telegram*, *YouTube*, *Instagram* etc., as quais são usadas para interações diversas entre usuários da Libras, sobretudo surdos, e, com isso, ficam os registros em vídeo. A partir do surgimento dos cursos de Letras-Libras em todo o país, as pesquisas linguísticas envolvendo a Libras começam a ganhar mais espaço e, então, percebe-se a necessidade de dados linguísticos com qualidade e sistematicidade para fins de pesquisa. Surgem, assim, as propostas de elaboração de *corpora* de Libras, a exemplo do *Corpus* de Libras da Grande Florianópolis, em Santa Catarina.

No âmbito da Universidade Federal de Alagoas, em 2014, o projeto *Corpus* da Libras da Grande Maceió (projeto replicado da UFSC) consistiu em documentar a Libras usada pelos surdos adultos da Grande Maceió. O projeto contou com a participação de bolsistas do Curso de Letras-Libras, dentre os quais, o autor dessa pesquisa. A dissertação, portanto, surge como repercussão de nossa experiência em documentação da Libras, desde 2014, quando participante do Programa Institucional de Bolsas de Iniciação Científica (PIBIC) até 2018, em que, por quatro anos, fomos confrontados com o desafio de coletar dados da Libras, editá-los e, principalmente, transcrevê-los.

A escolha do tema se deu quando comecei a estudar o Curso de Letras-Libras na Universidade Federal de Alagoas – UFAL, em 2014. Docentes do Curso, em visita técnica ao Letras-Libras da Universidade Federal de Santa Catarina – UFSC, tiveram acesso ao *Corpus* de Libras, projeto ali desenvolvido sob a coordenação da Professora Ronice Müller de Quadros. Aos moldes do projeto da UFSC, o Professor Jair Barbosa da Silva assumiu a coordenação do *Corpus* de Libras na UFAL. Como trabalho há quatro anos na minha experiência pela coleta de dados da *Corpus* de Libras da Grande Maceió, onde fui desafiado a pesquisar mais profundamente sobre esse tema, sabendo que o registro de um *Corpus* de Libras também auxilia em outras linhas teóricas de pesquisas linguísticas envolvendo a Libras. É muito importante que se tenha um *Corpus* de Libras representativo da Grande Maceió, uma vez que neste Estado não se dispõe de registros sistematizados desta Língua para pesquisas.

Em Maceió, os critérios de escolha dos participantes, registro, edição, transcrição, e tudo o que dispõe o processo de desenvolvimento do *corpus*, foi seguindo o modelo do *Corpus* de Libras da UFSC.

É justamente nesse contexto, sobre o processo de transcrição da variedade da Libras da grande Maceió, que surgem as nossas inquietações acerca das dificuldades e problemas para a transcrição e anotação de uma língua de sinais, para a qual ainda inexiste um sistema de escrita amplamente consistente e metodologias/estratégias de anotação que possam, de fato, recobrir a complexidade estrutural da Libras.

## **1.1 Objetivos da pesquisa**

### **1.1.1 Objetivos Geral**

- Compreender o processo da transcrição de dados do *Corpus* da Libras.

### **1.1.2 Objetivos Específicos**

- Discutir os desafios durante o processo de transcrição de dados linguísticos da Libras;
- Categorizar quais problemas podem estar relacionados às dificuldades com a transcrição;
- Propor soluções para transcrever os dados linguísticos da identificação e registro da variedade da Libras no *Corpus* de forma consistente e padronizada

## **1.2 Pergunta de pesquisa**

- a) Quais os entraves e possíveis soluções para os problemas encontrados no momento das transcrições da Libras com uso de glosas?

## **1.3 O problema da pesquisa**

No ano de 2002, foi sancionada a Lei Nº 10.436, que dispõe sobre a Língua Brasileira de Sinais – Libras, como forma de comunicação e expressão, de sistema linguístico com estrutura gramatical própria, oriunda de comunidades de pessoas surdas do

Brasil. Três anos depois, foi aprovado o Decreto Nº 5.626/2005, que regulamenta a Lei de Libras e, de forma mais ampla, este documento estabelece diretrizes legais para a implementação de um planejamento linguístico envolvendo os surdos e a Libras em diversas instâncias sociais, dentre elas, a educacional.

Como consequência disso, diversos cursos de Letras-Libras foram criados no Brasil, cerca de 40, e isso acaba impactando sobre a formação do pesquisador e da pesquisa envolvendo a Libras. Esta língua que teve suas primeiras pesquisas descritivas no Brasil, de forma sistemática e institucionalizada, no final da década de 1980, agora, por uma questão de política pública e linguística, passa a gozar de vasto espaço na academia. Em função disso, surge a necessidade de se conhecer a língua em termos linguísticos, o que demanda a sua documentação para posterior descrição.

Diferentemente da história das línguas orais, que mesmo antes dos gravadores, poderiam (e foram) registradas/documentadas por meio da escrita, as chamadas Línguas de Sinais, até recentemente, não podiam ser documentadas por falta de recursos tecnológicos adequados para este fim, inclusive pela ausência de um sistema de escrita de sinais, que segundo Estelita (2008), a ausência de usos sociais de um sistema de escrita de sinais, talvez, seja por uma imposição das línguas orais sobre as de sinais, hipótese não confirmada. Neste sentido, se quisermos saber como era a Libras usada no Brasil há 10 anos, certamente teremos bastante dificuldade pela ausência de *corpora* que tenham documentado esta língua. Isso traz consequências bastante relevantes para a Linguística Brasileira em termos de desconhecimento de uma língua brasileira ainda pouco conhecida pela Ciência e, em termos sócio-histórico-culturais, podemos mesmo dizer que há um apagamento do que foi a língua em tempos passados, mesmo que não haja um espaço de tempo tão significativo (20, 30 anos passados).

De acordo com Himmelmann (2006) *apud* Oliveira Jr (2013, p. 13), "a documentação linguística é a área que se ocupa com métodos, ferramentas e bases teóricas para a elaboração de um registro de uma língua natural, ou de uma variedade, que seja representativo, duradouro e que permita múltiplos usos". É justamente dessa dimensão científica (métodos, ferramentas e bases teóricas) de que fala Himmelmann que padece a Libras. Com o advento das novas tecnologias de comunicação, sobretudo das redes sociais, a Libras, no século XXI, circula de Norte a Sul fortemente no país, o que a torna mais viva, mas, ainda assim, em conformidade com os parâmetros estabelecidos pela UNESCO, trata-se de uma língua de sinais em risco de extinção, o que justifica, também, a sua documentação. Assim, se por um lado existe a necessidade de a Libras ser devidamente

descrita pela Linguística, a partir de dados que lhes sejam representativos e, portanto, metodologicamente bem recolhidos, por outro, é igualmente necessária a documentação desta Língua já que se trata de uma língua em risco nos termos tratados pela UNESCO. É importante notar, no entanto, que quando se fala em risco de extinção de uma língua, há uma gama de aspectos a serem observados. Conforme Leite e Quadros (2014), existe uma diferença entre o risco que as línguas de sinais nativas brasileiras correm e o risco da língua de sinais nacional, a Libras. No primeiro caso, de fato, há risco de extinção; já no segundo, parece mais adequado falar em *línguas em risco esquecidas*, (NONAKA, 2004) *apud* LEITE e QUADROS, 2014).

Diante desse contexto, o problema principal, para o qual se volta esta pesquisa, estão as inúmeras dificuldades com as quais os transcritores se deparam no momento da transcrição dos dados. Algumas destas dificuldades são comuns a trabalhos de transcrição de qualquer língua, como a segmentação e a busca por equivalentes linguísticos, quando a transcrição é feita em outra língua, no entanto, em se tratando de *corpus* de línguas de sinais, as dificuldades são ainda maiores, pois a glosagem dos sinais é feita numa língua de modalidade diferente, o que gera muitos percalços neste processo.

#### **1.4 Estrutura da dissertação**

A dissertação encontra-se estruturada em uma Introdução, em que apresentamos o objetivos, justificativa e problemas da pesquisa, mais três capítulos: no segundo capítulo, são apresentadas as bases teóricas para a pesquisa; no terceiro, descreve-se o percurso metodológico do trabalho e, por fim, no quarto capítulo, apresenta-se a análise dos dados. Finalizando o trabalho, têm-se as considerações finais e as referências.

## **II. CAPÍTULO – Fundamentação teórica**

Neste capítulo, serão apresentados os fundamentos da Linguística de *Corpus* referentes à documentação de línguas, especialmente das línguas de sinais, bem como os processos a serem realizados que são inerentes à metodologia utilizada para a coleta de dados no *Corpus* de Libras. Além disso, apresentaremos como é descrito em um manual de orientação para a construção de *corpus* de Libras o processo transcrição dos dados para a

Libras utilizando glosas de um *SignBank* específico. Esta descrição nos ajudará a entender as dificuldades encontradas nesta pesquisa.

## 2.1 A Linguística de *Corpus*

Segundo Hereweghe e Vermeerberge (2012), a Linguística de *Corpus* é um ramo relativamente novo da pesquisa linguística, o qual anda de mãos dadas com as possibilidades oferecidas pelos recursos tecnológicos cada vez mais avançados no século XXI. No passado, qualquer conjunto de dados por meio do qual uma análise linguística era realizada denominava-se de *corpus*. No entanto, com o advento da informática e da linguística baseada em *corpus*, o uso do termo *corpus* tornou-se restrito a qualquer tipo de coleção de textos em formato legível/processado por máquina (os computadores).

Para Johnston (2009, p. 18) *apud* Hereweghe e Vermeerberge (2012, p. 1033), a linguística do *corpus* baseia-se no pressuposto de que o processamento de grandes quantidades de textos anotados pode revelar padrões de uso e estrutura da linguagem não disponíveis por dados de intuição ou mesmo por uma análise linguística de um texto particular, o que implica, necessariamente, uma organização teórico-metodológica em termos de coleta, tratamento e armazenamento dos dados.

Já para Sardinha (2000),

a Linguística de *Corpus* ocupa-se da coleta e exploração de *corpora*, ou conjuntos de dados linguísticos textuais que foram coletados criteriosamente com o propósito de servirem para a pesquisa de uma língua ou variedade linguística. Como tal, dedica-se à exploração da linguagem através de evidências empíricas, extraídas por meio de computador (SARDINHA, 2000).

Este autor argumenta que, embora hoje a Linguística de *Corpus* esteja diretamente vinculada ao computador, à possibilidade de armazenamento e manipulação de grande quantidade de dados via computador, "havia *corpora* antes do computador, já que o sentido original da palavra '*corpus*' é 'corpo', 'conjunto de documentos' (conforme o dicionário Aurélio)". Desde a Antiguidade, *corpora* eram produzidos a partir de citações das Escrituras Sagradas (Bíblia).

Ainda segundo Sardinha (Op. Cit), na área de Linguística, um dos *corpora* linguístico eletrônico mais antigo de que se tem conhecimento é o *corpus Brown*, o qual foi lançado em 1964 e, para os padrões da época, continha uma quantidade invejável de dados: um milhão de palavras. O *corpus Brown* constitui um marco importante para o que hoje se chama de Linguística de *Corpus*. Sua importância se dá não só por questões metodológicas,

mas também por uma razão histórica: o *corpus Brown* surge sete anos após o lançamento de "*Syntactic Structures*", obra publicada por Chomsky, que coloca em discussão a própria ideia de *corpus*, uma vez que nesse novo paradigma da Linguística os dados de que o linguista precisaria para suas análises estão todos em sua mente, acessíveis por meio da introspecção, portanto, a ideia de se fazer pesquisa com uso de dados provenientes de *corpora* estava desacreditada e havia mesmo hostilidade para essa "velha" perspectiva.

Para Fenlon, Schembri, Johnston e Cormier (2015, p. 157), um *Corpus* linguístico é uma base de referência por meio de uma catalogação e registro de idiomas falados (Línguas orais), escritos e sinalizados (Línguas de Sinais), que tem metadados agregados, sendo legíveis por computador, e que na medida do possível representam a língua em uso dos seus falantes nativos, permitindo um estudo sobre a frequência de alguma palavra ou sinal ou tipo de construções nessa língua. Dessa forma, o uso do *software ELAN*<sup>1</sup>, por exemplo, pode ser uma estratégia ideal para a identificação e registro das variedades que uma determinada Língua de Sinais em foco de pesquisa apresenta no decorrer do tempo, uma vez que esse recurso tecnológico possibilita a transcrição da língua sincronizada com textos em vídeos.

Também com base em McCarthy e O'Keefe (2010, p. 7), a Linguística de *Corpus*, para muitos, é "um fim em si mesma", ou seja, ela permite que de forma empírica a língua em foco seja analisada, definida e descrita. A autonomia, contribuição e autenticidade de um *Corpus* pode impactar em várias áreas como por exemplo: na área da linguística forense, análise de discurso, pragmática, dentre outras. Para além de uma perspectiva da ciência linguística, pesquisadores de outras vertentes e abordagens podem utilizar dados obtidos pela Linguística de *Corpus*, especialmente na área dos estudos das Línguas de Sinais cujos estudos, em comparação com outras línguas de modalidade oral, ainda são bastante insólitos.

### 2.1.1 A Sociolinguística

Como já explicitado acima, em grande parte, os *corpora* têm por finalidade, embora não exclusivamente, analisar as variações sociolinguísticas de uma dada língua. Neste sentido, Linguística de *Corpus* e Sociolinguística, especialmente a quantitativa, dialogam de modo bastante proximal. Ambas lidam com grande quantidade de dados, embora para fins

---

<sup>1</sup> *ELAN (Eudico Linguistic Anotator)* é um *software* largamente usado por linguistas para fazer anotação e transcrição de línguas orais e de sinais, o qual permite sincronizar anotações com áudio e vídeos, além de possibilitar a interface com outros *softwares* e permitir a busca de dados anotados.

distintos, o que implica a necessidade de uma metodologia específica e teoricamente fundamentada para que os dados coletados sejam comparáveis, úteis para a pesquisa e tenham qualidade, sem o que torna inviável o uso para pesquisas futuras. Segundo Coelho, Gorski, Souza, May, (2015, p.14), "os pesquisadores dessa área (Sociolinguística Quantitativa) costumam lidar com uma grande quantidade de dados de usos da língua, o que requer normalmente uma análise estatística".

Conforme Cezario e Votre (2016, p. 141),

a Sociolinguística é uma área que estuda a língua em seu uso real, levando em consideração as relações entre a estrutura linguística e os aspectos sociais e culturais da produção linguística. Para essa corrente, a língua é uma instituição social e, portanto, não pode ser estudada como uma estrutura autônoma, independente do contexto situacional, da cultura e da história das pessoas que a utilizam como meio de comunicação.

Para estes autores, a Sociolinguística toma como pressuposto básico o fato de que variação e mudança linguísticas fazem parte de qualquer língua, razão por que devem ser consideradas em qualquer análise linguística. É neste sentido que o sociolinguista "se interessa por todas as manifestações verbais nas diferentes variedades de uma língua" (CESÁRIO e VOTRE, 2016, p. 141).

A Sociolinguística atua na interface língua/sociedade, o que significa dizer que os fatores sociais condicionam/interferem nas estruturas linguísticas. "São muitas as áreas de interesse da Sociolinguística: contato entre as línguas, questões relativas ao surgimento e extinção linguística, multilinguismo, variação e mudança constituem temas de investigação na área" (MOLLICA, 2015, p. 10). Já para Bagno (2007), a noção de variação linguística constitui a "espinha dorsal" da Sociolinguística. Para este autor, a ideia de considerar a variação linguística implica, necessariamente, dizer que a língua é heterogênea e isto não significa que as línguas são um "caos"; "mesmo apresentando diferenças, elas se entendem perfeitamente, e que isso é possível porque a língua é um sistema organizado, formando por regras categóricas e regras variáveis" (COELHO, GORSKI, SOUZA e MAY, 2015, p. 59). Argumentam ainda as autoras:

A língua que falamos nos põe à disposição diferentes formas para expressar os mesmos significados, sem perder sua sistematicidade ou seu poder como instrumento de comunicação. Apontamos também que a Sociolinguística é a subárea específica da Linguística com teoria e métodos voltados à compreensão da variação e a mudança nas línguas (COELHO, GORSKI, SOUZA e MAY, 2015, p. 55).

Considerando que a variação e a mudança linguísticas são inerentes a qualquer língua, isso não poderia ser diferente com as línguas de sinais. Segundo Quadros e Karnopp

(2004) a "compreensão da Libras como língua natural, com gramática própria e que precisa ser estudada em seus aspectos fonológicos e morfossintáticos", o que traz como consequência para o estudo observar a variação e a mudança linguísticas. Assim, faz-se necessária a constituição de *corpora* da Libras, teórico-metodologicamente consistentes, a fim de se ter uma quantidade de dados reais da língua suficientes para se lançar hipóteses sobre a língua, testá-las e traçar generalizações pertinentes acerca da estrutura e funcionamento dessa Língua.

Karnopp (s/d, p. 6-7) afirma que "ao estudarmos as Línguas de Sinais, estamos tratando também das relações entre linguagem e sociedade". Sendo a Libras a língua majoritariamente falada pela comunidade surda brasileira, ao se registrar e descrever esta língua se constata, de imediato, a diversidade de usos nessa comunidade linguística, a qual "se caracteriza pelo emprego de diferentes modos de usar a Língua de Sinais. As essas diferentes maneiras de fazer sinais, utiliza-se a denominação de variedades linguísticas" (KARNOPP,s/d, p. 6-7).

Como contribuição Nyst (2015, p. 108) explicita que no campo de trabalho de coleta de dados linguísticos é necessário que esse trabalho seja realizado em um "ambiente natural", ou seja, coletando a língua na sua produção espontânea sem manipulação e fingimentos, permitindo uma coleta e registro da língua em seu uso natural, preferencialmente fora de ambientes como laboratórios ou bibliotecas (observe em Bower, 2008; Majid, 2012). No caso da Libras, caso não haja uma sinalização natural, isso acarretará num prejuízo de dados verídicos, que de fato representem extratos da língua. A coleta de dados não se esgota em um registro único e absoluto, pois é possível que, após 10 anos, novas coletas mostrem um comportamento linguístico diferente por parte dos usuários. Neste sentido, coletar dados de uma língua em um dado momento pode ser muito útil para realizar análises diacrônicas da língua.

Na perspectiva de Fenlon, Schembri, Johnston e Cormier (2015, p. 157), os linguistas que se debruçam nos estudos das línguas de sinais se referem ao termo "*corpus*", como qualquer agrupamento de dados que uma descrição ou análise linguística tem, ou a qualquer coleção de gravações em vídeos da língua em uso (Lucas, Bayley e Valli, 2001; Neidle e Vogler, 2012). E percebe-se que esses agrupamentos e coleções que equivalem a *corpora* são de muita valia, pois ajudam os pesquisadores a terem um embasamento melhor nas suas pesquisas envolvendo *corpora* de diversos tipos, sempre tendo como referência surdos nativos da Língua de Sinais em suas análises linguísticas.

A coleta de dados naturalísticos é um desafio caro à Sociolinguística Variacionista, tendo Labov como um dos mais importantes pesquisadores que se debruçou sobre essa questão desde a década de 70. No que concerne às línguas de sinais, um grande gargalo: por um lado necessitamos de um vasto aparato tecnológico (câmera, luz, tripés, estúdio) para garantir dados com qualidade, o que, por outro lado, impacta na naturalidade dos dados.

### 2.1.2 Construção de *corpus* e tecnologias

McCarthy e O' Keeffe (2010, p. 5) falam que o que permitiu que a atual Linguística de *Corpus* que conhecemos hoje pudesse ganhar consistência em termos tecnológicos foi o desenvolvimento de *Hardware* e *Software* entre as décadas de 1980 e 1990. Havia a necessidade de tecnologias serem desenvolvidas a fim de se manipular e registrar grandes quantidades de dados para os primeiros registros modernos de *corpus* serem efetuados. Sendo assim, os *softwares* auxiliam grandemente nessa possibilidade, valorizando as Línguas de Sinais, conforme Nonaka, (2004), possibilitando que além de línguas orais, as Línguas de Sinais também tenham seu registro no decorrer dos séculos. No ano de 1990, o surgimento do primeiro desktop revolucionou os trabalhos dos grupos de pesquisa com *corpora*, de modo que se deu o pontapé inicial nessas pesquisas, sem haver a necessidade de se preocuparem sobre os termos de “autorização”, que já tinham à disposição as imagens de cada pessoa em registro. Ainda numa perspectiva tecnológica, é importante que se tenha um espaço de armazenamento com melhor qualidade possível, pois sabe-se que a Libras é uma língua com modalidade visuogestual (fato aplicável a qualquer Língua de Sinais) e que isso exige uma captação com rigorosidade científica para sua melhor visualização, além de horas ilimitadas para um maior conforto e liberdade temporal no momento da entrevista, fazendo-se necessário um armazenamento maior e suficiente.

Fenlon, Schembri, Johnston e Cormier (2015, p. 158) destacam que até pouco tempo atrás não era possível registrar por meio de *Corpora* as Línguas de Sinais, uma vez que apenas com o avanço tecnológico recente é que se tornou possível o registro dessas línguas por meio de vídeos. O desenvolvimento de recursos tecnológicos permite rápido processamento e armazenamento das línguas de sinais, com o acarretamento de mais precisão nas análises dos dados. Parece imperativa a ideia de que para a investigação em linguística de línguas de sinais, a tecnologia é elemento fundamental.

Nessa mesma linha, McCarthy e O' Keeffe (2010, p. 6) afirmam que o principal fator que potencializa o crescimento e desenvolvimento da Linguística de *Corpus* é a tecnologia. Através dela tornou-se possível fazer armazenamento de grandes quantidades de dados coletados em computadores pequenos, por exemplo, sendo possível fazer o registro de um *corpus* do tamanho que seja desejado. Isso é visto na grande quantidade de dados (vídeos de alta qualidade na resolução), onde é preciso tecnologia de última geração para que a captura de imagens detalhadas não sejam perdidas, mas sejam salvaguardadas para futuras pesquisas e análises. Exemplo disso é o portal da UFSC, onde há o armazenamento de dados da Libras para que em seguida possam ser disponibilizados ao acesso do público - <https://corpuslibras.ufsc.br/>. Ainda a título de ilustração, algumas ferramentas de ordem tecnológicas são amplamente usadas para a elaboração de *corpora* em Línguas de Sinais, a saber: figura 1, o *Adobe Premiere* para edição de vídeo; figura 2, o *software ELAN*, para anotações e transcrições de dados; figura 3, o *SignBank*, para busca de itens lexicais para as anotações e, por outro lado, catalogação de sinais inexistentes no *SignBank*. Veja abaixo:

**Figura 1 - Adobe Premiere Pro - Edição de vídeo**



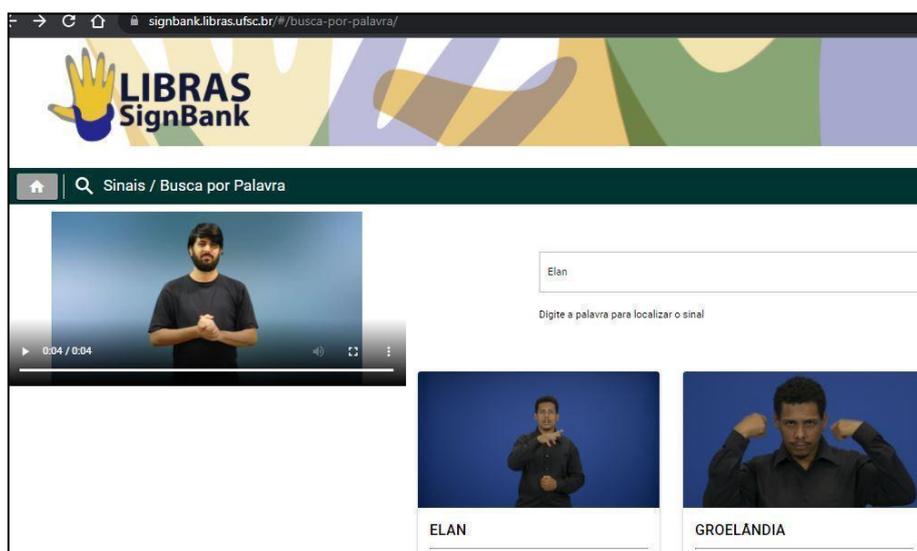
Fonte: *Adobe Premiere Pro*. Imagem disponível em: <<https://www.adobe.com/br/products/premiere.html>>. Acesso em: 30 de junho de 2022.

**Figura 2 - Software ELAN - Linguistic Annotator**



Fonte: *Software ELAN*. Imagem disponível em: <https://archive.mpi.nl/tla/elan/download>. Acesso em: 30 de junho de 2022.

**Figura 3 - LIBRAS SignBank**



Fonte: *SignBank - Libras*. Imagem disponível em: <https://signbank.libras.ufsc.br/#/>. Acesso em: 30 de junho de 2022.

## 2.2 Elaboração de *corpora* em Línguas de Sinais

Na área de Línguas de Sinais, o *corpus* linguístico (pelo menos em seu sentido mais restrito de *corpus*) ainda está em fase bastante inicial; pode-se dizer que em sua infância, embora crescendo rapidamente em diversos países. Para Johnston (2008, p. 82), *apud* Hereweghe e Vermeerberge (2012), a necessidade de *corpora* em Língua de Sinais pode ser

esboçada nos seguintes termos: os *corpora* em Línguas de Sinais são de extrema importância para as pesquisas linguísticas envolvendo estas línguas, em todos os níveis linguísticos, da fonologia ao discurso, sobretudo porque falta às Línguas de Sinais um sistema de transcrição que não seja baseado em Línguas Orais, além do estatuto linguístico de línguas minoritárias do qual gozam as línguas sinalizadas. Desta forma, embora dados provenientes da introspecção e da observação possam revelar hipóteses sobre o uso e estrutura das Línguas de Sinais, pesquisas com base nesse tipo de abordagem podem apresentar falhas e mesmo dissenso quanto a certas questões de natureza gramatical, colocando em "xeque" a aceitabilidade de determinadas generalizações sobre a língua.

Para Hereweghe e Vermeerberge (2012, p. 1035),

Um dos primeiros (se não o primeiro) projeto de *corpus* de Língua de Sinais em grande escala é o *corpus* da *American Sign Language (ASL)* coletado por Ceil Lucas, Robert Bayley, e seus colaboradores. No decorrer de 1995, eles coletaram dados em sete cidades nos Estados Unidos que foram considerados representativos das principais áreas do país: Staunton, Virginia; Frederick, Maryland; Boston, Massachusetts; Olathe, Kansas / Kansas City, Missouri; Nova Orleans, Louisiana; Fremont, Califórnia; e Bellingham, Washington.

O critério para a coleta de dados nessas cidades é justamente o fato de elas terem comunidades prósperas usuárias de *ASL*, bem como algumas residências escolares para crianças surdas. Participaram deste *corpus* da *ASL* 207 afro-americanos, trabalhadores brancos, homens e mulheres de classe média, os quais foram subdivididos em três grupos etários: 15 - 25, 26 - 54 e, por fim, acima de 55 anos. Todos os informantes adquiriram *ASL* como língua nativa em casa ou nas residências escolares antes dos 5 ou 6 anos de idade.

De 1995, quando do surgimento do *corpus da ASL*, para os dias atuais diversos projetos envolvendo Línguas de Sinais foram empreendidos: Austrália, Irlanda, Holanda, Reino Unido, Alemanha, China (Hong Kong), Itália, Suécia e França. Mais recentemente, países como Espanha, Portugal e Brasil têm investido na elaboração de *corpora* de suas respectivas Línguas de Sinais. Grande parte desses *corpora* tem como foco questões relativas à variação sociolinguística; outros, no entanto, atendem a objetivos múltiplos, como preservação das línguas, ensino de línguas, descrição linguística e artefato cultural.

Pensando ainda do porquê temos a necessidade de se ter *corpora* de Línguas de Sinais, temos essa indagação respondida pelos autores Fenlon, Schembri, Johnston e Cormier (2015, p. 157-158):

Por que precisamos de *corpora* de Língua de Sinais? Uma razão importante é que há muito trabalho a ser feito no campo da linguística da Língua de Sinais para aprofundar nossa compreensão da estrutura e do uso dessas línguas. Há uma necessidade urgente para testar as afirmações feitas por muitas descrições e análises linguísticas existentes de Línguas de Sinais, porque elas frequentemente

se baseiam em conjuntos de dados limitados de um pequeno número de sinalizantes<sup>2</sup>. (tradução nossa)

Mesmo com as melhores das intenções e bastante esforço no desenvolvimento de *corpora* de Línguas de Sinais, é preciso lembrar que ainda há muito a ser pesquisado e sabemos que a pesquisa a partir de *corpus* da Língua de Sinais ainda está engatinhando, e se desenvolvendo com o passar do tempo, o que terá como consequência mais consistência e o aprimoramento em termos de recursos tecnológicos e metodológico para o desenvolvimento de pesquisas envolvendo *corpora* com essas línguas.

Para McCarthy e O'Keeffe (2010, p. 4) é notável a importância da coleta de dados e que esses dados sejam verídicos, ou seja, as sinalizações em foco de registro (no caso das Línguas de Sinais) precisam ser espontâneas. Com um delicado trabalho de *corpus*, refletido na construção de um dicionário, o Dr. Samuel Johnson desenvolveu coleta de dados do inglês, a qual foi registrada em longas e longas tiras de papel, no período de 1560 até 1660, até que em 1755 teve sua publicação. Isso evidencia dois fatos importantes: a dinamicidade dos objetivos de um *corpus* linguístico, que ora se presta a apenas descrever uma língua, ora se presta à elaboração de dicionários, por exemplo, e, por outro lado, a escassez de recursos tecnológicos à época do Dr. Samuel Johnson, em que para se coletar dados o uso do papel e da escrita eram os recursos de que se dispunha.

Com o passar do tempo, a tecnologia se desenvolveu e o *Corpus* também acompanhou esse desenvolvimento, tornando o registro linguístico mais prático e rápido, não sendo mais necessário metros e mais metros de papel, nem horas e horas de anotações à mão, além de ser possível fazer registros de áudios das línguas orais, pois o *corpus* pioneiro não tinha essa possibilidade.

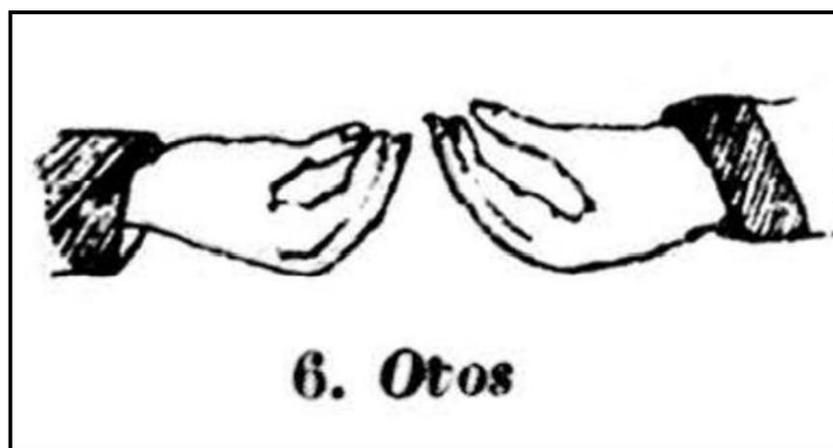
De modo similar, quando focamos no registro precursor de Sinais, percebemos que demandava muito mais tempo e esforço catalogar e registrar os sinais por meio de desenhos como fez Gama, dando início no registro de Sinais em 1875, o que quase sempre ocorria por meio de desenhos seguidos de alguma explicação sobre desenho por meio escrito. Atualmente, com o avanço da tecnologia, é possível fazer registros em vídeos de alta definição e excelente qualidade da Língua de Sinais. Pesquisadores mais antigos recorriam ao uso de desenhos a fim de mostrarem como as línguas de sinais funcionavam, o que

---

<sup>2</sup> Why do we need sign language corpora? One important reason is that there is much work that needs to be done within the field of sign language linguistics to further our understanding of the structure and use of these languages. There is a pressing need to test the claims made by many existing linguistic descriptions and analyses of sign languages, because they have often been based on limited data sets from a small number of signers.

certamente limitava as análises. Em função dessa limitação da imagem parada, por meio de desenho, os autores têm a necessidade de descrever com detalhes o que querem mostrar com as imagens, incluindo-se aí não apenas as configurações de mãos, mas também os possíveis movimentos realizados quando da execução de um dado sinal. Na contemporaneidade, os recursos tecnológicos auxiliam nas transcrições e minimizam os impactos negativos dos desenhos. Veja-se, a título de exemplificação, como Gama (1857) representa OVOS.

**Figura 4 - OVOS**



**Fig. 6.** — Bater com as extremidades dos dedos, umas contra as outras, como se se quebrasse ovos batendo um no outro.

Fonte: Gama (1857, p. 14 e 15).

Com suas particularidades culturais e linguísticas, a realização de um *corpus* começa a ser constituída a partir da própria vivência do pesquisador com a língua em foco, pois um *corpus* da Língua de Sinais é mais do que apenas sinais aleatórios registrados por vídeo. De maneira sensível, o pesquisador também precisa discernir em quais contextos os léxicos anteriores ao registro se encaixam.

Em relação ao primeiro projeto de *corpus* em Língua de Sinais, que começou em 2004 na Austrália, com arquivos digitais de vídeo, foram registradas as produções de 100 surdos nativos da *Auslan* (*Australian Sign Language*) e sinalizantes que adquiriram a Língua Australiana de Sinais cedo (Fenlon, Schembri, Johnston e Cormier 2015, p. 160), nos remetendo à importância do paralelo que existia entre o desenvolvimento da tecnologia e o uso das novas possibilidades disponíveis.

Nonaka (2004) direciona para as Línguas de Sinais uma visão de risco de serem perdidas ou apagadas da história, pois suas discussões remetem ao fato de que até os trabalhos de William Stokoe em 1960 serem publicados, evidenciando para a linguística que

a *ASL* (*American Sign Language*) era de fato uma língua e que os surdos não se comunicavam de forma aleatória e sim de maneira estruturada, utilizando uma língua propriamente dita, se percebe que antes disso resultou numa privação de herança para as comunidades surdas não só usuárias de *ASL*, como também a todas as Línguas de Sinais existentes. Concordando com o autor citado, urge que a catalogação e pesquisa em registros da Língua de Sinais continuem se desenvolvendo, pois os registros são patrimônio da comunidade surda. As línguas são vivas e não se ter registro das gerações passadas para fins de análise linguística e apreciação implica prejuízos histórico-culturais e linguísticos. Assim como os registros de línguas arcaicas que eram esculpidos e desenhados em pedras, e que atualmente são usados para ensinar as novas gerações aspectos culturais das primeiras civilizações, perpetuando sua história como patrimônio da humanidade, da mesma forma é necessário que exista um olhar científico para o registro das Línguas de Sinais como forma de respeito, perpetuação e valorização da cultura surda, a qual é minoria linguística.

Ainda com base em Fenlon, Schembri, Johnston e Cormier (2015, p. 158), à medida que o tempo passa, os *corpora* são um meio seguro e de grande valia para os registros de Línguas de Sinais, pois as línguas são vivas, naturalmente se modificam por meio da globalização, avanços tecnológicos e dos próprios nativos. Por meio de *corpora* as línguas são perpetuadas, assim como acontecimentos nas nossas vidas são guardados na memória. Também é interessante que essas línguas de sinais minoritárias estão em risco de extinção, nos remetendo à grande urgência de continuarem a ser catalogadas e individualmente coletadas (JOHNSTON, 2004; NONAKA, 2004; SCHEMBRI, 2010).

Ainda nessa direção Singleton, Martin e Morgan (2015, p. 12) declaram que nos últimos 10 anos, foi percebido um aumento na quantidade de projetos de pesquisas em Línguas de Sinais, não só das Línguas de Sinais majoritárias, mas também estudos em Línguas de Sinais emergentes e de minoria linguística em todo o mundo. Isso reforça o fato de também ser necessário mais publicações e pesquisas na catalogação e registro das Línguas de Sinais Indígenas e minoritárias (como a Língua de Sinais Cena e Urubu-Kaapor no Brasil), que também estão à disposição de novos pesquisadores e novas possibilidades de contribuição na área da linguística. Tendo particularidades culturais e linguísticas tão ricas quanto outras Línguas de Sinais com maior número de sinalizantes, além de obviamente ter o mesmo status linguístico das línguas orais, afinal língua é língua, independentemente de ser sinalizada ou falada.

Mas e a respeito da equipe técnica responsável pela manipulação de um *corpus* da Língua de Sinais, como é na prática? De acordo com

Singleton, Martin e Morgan (2015, p. 15), os pesquisadores na dianteira da pesquisa devem escolher e treinar sua equipe de pesquisa com cuidado. No mínimo, os membros da equipe de pesquisa devem ter fluência básica numa Língua de Sinais e conhecimento cirúrgico das questões centrais que se remetem à cultura das comunidades surdas (embora reconheçam que nem todas as comunidades surdas compartilham os mesmos valores)<sup>3</sup>. (tradução nossa)

É delicado e cirúrgico o trabalho do pesquisador no registro de uma língua por meio de um *Corpus*. A equipe precisa ter não somente conhecimento técnico, mas também entender o básico de uma Língua de Sinais, além de mergulhar na cultura e ter intimidade com as questões políticas e sociais do grupo em registro, para que suas nuances culturais sejam internalizadas, propiciando um melhor desempenho nos registros dos dados. Daí, são iniciados os trabalhos de captação por vídeo, com rigor científico, em seguida são feitas as edições dos vídeos, e seguem-se os passos no desenvolvimento do *corpus* da Língua de Sinais.

Com base em Fenlon, Schembri, Johnston e Cormier, (2015, p. 159), sabemos que ainda existe muita pesquisa a ser feita e desenvolvida, sendo de suma importância que o *Software ELAN* também continue sendo aprimorado, possibilitando um melhor trabalho em *Corpora* de Língua de Sinais (Crasborn e Sloetjes, 2008), principalmente na etapa de anotação, que é bastante demorada e trabalhosa de se realizar, facilitando também o desenvolvimento e criação de um *SignBank* mais completo e atualizado. E por meio do registro em vídeo de alguns convidados, sinais que ainda não estão no *SignBank*, serão consequentemente inseridos no *SignBank*.

Quando nos direcionamos ao que disseram Johnston e Schembri (2007) e Sutton-Spence e Woll (1999), notamos que um projeto de *corpus* linguístico que engaje participantes nativos da Língua de Sinais em questão, que sejam naturais do país, mas que não sejam naturais da mesma região, é bastante proveitoso, uma vez que a Língua de Sinais tem variações e esses participantes de regiões diversas possibilitarão a que essas diferenças/variações linguísticas apareçam no *corpus* da Língua de Sinais, mostrando a diversidade linguística do grupo de sinalizantes eleitos para representar o *corpus*.

---

<sup>3</sup> Principal investigators should choose and train their research team carefully. At a minimum, research team members should have basic fluency in a world sign language and knowledge of core issues in Deaf cultures and communities (while acknowledging that not all Deaf communities share the same values).

## 2.3 *Corpus* de Libras

### 2.3.1 A Documentação da Libras

A documentação de uma língua passa, necessariamente, por um planejamento que vai além dos aspectos linguísticos. Tradicionalmente, desde os tempos mais remotos (Antiguidade), se tem conhecimento de registros de línguas orais; os sistemas de escrita e mais tarde os gravadores de áudio sempre favoreceram a documentação de línguas cuja produção se dá por meio da emissão de voz. Quando nos atemos às Línguas de Sinais, apenas por meio dos registros em vídeo é que se pode captar essas línguas, dada a sua modalidade de produção, que é visuogestual. Mesmo com o surgimento do aparato tecnológico capaz de captar a produção de fala dos sinalizantes, pouco há registrado de 30 anos para trás, por, basicamente, dois motivos: i) historicamente, por longa data, os surdos foram proibidos de usarem as Línguas de Sinais; ii) até a década de 1990, os dispositivos tecnológicos para armazenamento de vídeo eram extremamente perecíveis.

De modo semelhante ao que ocorreu com o restante do mundo, é na década de 1990 que começa a ganhar espaço a documentação da Libras. Para Quadros (2016, p.11),

O *Corpus* da Língua Brasileira de Sinais (Libras) começou a ser constituído em 1995. Este *Corpus* envolve diferentes projetos, compreendendo dados de fontes diversas e diretrizes para o registro dos dados e metadados em Libras. O primeiro deles documenta dados de estudos longitudinais com crianças surdas filhas de pais surdos adquirindo a Libras. Posteriormente, foram incluídas crianças surdas filhas de pais ouvintes; crianças surdas com implante coclear, filhas de pais surdos e pais ouvintes; e crianças ouvintes filhas de pais surdos (Codas).

Documentar uma língua constitui uma tarefa que, para além da possibilidade de descrição e análise linguística, configura-se como o registro de um patrimônio imaterial de uma dada comunidade de fala, nos termos do Inventário Nacional da Diversidade Linguística (INDL), o qual foi instituído pelo Decreto N° 7.387, de 09 de dezembro de 2010. Dessa forma, trata-se de uma atividade de suma importância para a academia, mas, sobretudo, para o registro da Cultura e da História de um povo, no caso, do povo surdo.

Como em Quadros, Leite, Lohn, Schmitt e Pego (2020, p. 1), vemos que em 1875 Gama inicia a documentação da Língua Brasileira de Sinais tendo como base um documento que na época já era utilizado e que foi produzido por Pierre Pélissier, na França, o que resultou na publicação intitulada: “Iconographia dos Signaes dos Surdos-Mudos”, que inclusive está contida nos arquivos do INES (Instituto Nacional de Educação de Surdos), onde se diferencia da publicação original de Pierre Pélissier por causa das glosas que estavam em português e não em francês, como língua da publicação modelo. Após essa

produção, somente depois de 1969 é que a Língua de Sinais voltou a ter registro novamente, quando os trabalhos pelo missionário Eugênio Oates foram iniciados. Assim, percebemos que entre o período de 1875 a 1969 há uma lacuna histórica e uma perda de patrimônio linguístico irreparável. É interessante notar uma grande similaridade entre a Libras e a Língua de Sinais Francesa (LSF) nas primeiras documentações realizadas, com base na publicação de Campello (2009), de forma que percebemos que o registro de uma língua, também nos permite ver distanciamento históricos e linguísticos de outras línguas.

Daí, percebemos a importância de uma articulação que envolva o apoio de políticas públicas em prol da preservação e salvaguarda de línguas minoritárias na contemporaneidade. Assim, com base em Instituto do Patrimônio Histórico e Artístico Nacional - IPHAN (2014, apud Royer (2019, p. 78), vemos que

A política da diversidade linguística procura articular diferentes dimensões do Estado e da sociedade civil para a valorização e promoção das línguas minoritárias faladas no Brasil, vindo ao encontro do movimento crescente, em nível mundial, pela garantia de direitos linguísticos... os falantes têm o direito de usar suas línguas nos ambientes públicos e privados e de transmiti-las para as futuras gerações.

Ou seja, por trás do registro de uma língua, podemos ter acesso à Cultura, lutas, direitos e, com isso, analisar os níveis linguísticos da língua. Por outro lado, quando não se tem registro de uma língua, tudo isso é perdido, deixando uma grande lacuna na história do Povo que é nativo deste idioma. Lembrando que o registro linguístico não se limita a apenas análises linguísticas que envolvam agregar conteúdo à ciência (grande área da linguística), é importante também, que as próximas gerações possam ter o direito e a autonomia de apreciar o desenvolvimento temporal/cronológico da sua língua, bem como acesso às peculiaridades da sua herança linguística em épocas passadas. Com esse pressuposto, ressalta-se a importância da criação contínua e persistência de implementações de políticas públicas que envolvam a defesa da salvaguarda de uma língua.

No que tange à salvaguarda e ao inventariamento de Línguas de Sinais, destaque-se que:

No que se refere às Línguas de Sinais nativas e originais, o fato de estarem associadas a pequenas comunidades que vivem e sobrevivem sob o jugo de grandes estados-nações as coloca claramente numa situação de “risco”, bastante similar à de línguas orais faladas por outros grupos sociais minoritários (e.g. imigrantes e indígenas). (QUADROS e LEITE, 2014 *apud* ROYER, 2019, p. 79).

À medida que o tempo passa, a globalização também continua se infiltrando em várias frestas da sociedade e a Língua, por ser viva, continua evoluindo e se modificando, tendo influências da própria globalização, e como resultado disso as línguas naturalmente

podem ter influências umas sobre as outras. Porém, no caso dos surdos, e de outras comunidades que possuem uma língua minoritária, e vivem ao redor de comunidades muito maiores, linguisticamente falando, isso pode ser preocupante, pois é importante que as próximas gerações saibam de aspectos relacionados à sua cultura com base em registros da sua língua e em como ela se desenvolveu ao longo do tempo, além de garantir que essa língua não seja extinta. Com isso, é clara a necessidade de se registrar essas línguas minoritárias e de criar políticas públicas voltadas à preservação e salvaguarda das línguas que são patrimônios da humanidade.

Ainda sobre línguas minoritárias, em especial as Línguas de Sinais, Leite e Quadros (2014, p. 11) nos mostram que até hoje ainda existem Línguas de Sinais indígenas e que depois de um mapeamento, Vilhalva (2009; 2012) também pôde contribuir, analisando particularidades de surdos que usavam uma língua emergente de sinais para se comunicar. Mostrando também, que até mesmo comunidades minoritárias possuem alto potencial linguístico para desenvolver outras Línguas de Sinais e que em nível nacional temos muito ainda a descobrir sobre Línguas de Sinais emergentes, sendo bastante preocupante o fato de não se ter catalogação e salvaguarda dessas línguas em potencial, as quais apresentam o risco de serem apagadas da história humana, assim como muitas outras já o foram, deixando uma lacuna não só em termos científicos, mas, também, e não menos importante, saciar a curiosidade das novas gerações, em saber a respeito dessas línguas nos tempos passados, e de suas influências e desenvolvimento por meio da comunidade nativa.

Leite e Quadros (2014, p. 9) afirmam que ainda que se tem muito pouco conhecimento a respeito de outras variedades de Línguas de Sinais no Brasil, e que antes do INES ter sido atuante com o ensino da Libras no Brasil, também não temos nenhum documento sobre as Línguas de Sinais nativas do país. Mas pelo menos duas línguas (Urubu-kaapor e a Cena) já foram minimamente documentadas.

Atualmente o *Corpus* da Língua Brasileira de Sinais vem se desenvolvendo e mostrando resultados positivos, pois esses registros e pesquisas podem ser compartilhados com nativos de outras Línguas de Sinais aqui no Brasil, possibilitando que enxerguem uma amostra da riqueza de uma Língua de Sinais, de modo a influenciar o registro de outras Línguas de Sinais.. E à medida que mais pesquisadores publiquem artigos científicos, as Línguas de Sinais se tornarão mais conhecidas pela sociedade e melhor será para o reconhecimento e valorização dessa língua. Vemos que ainda existe muito a ser aprendido, principalmente quando pensamos nas Línguas de Sinais Indígenas. Quando paramos para analisar a história, lembramos que a Língua de Sinais já era existente no território brasileiro,

dando margem para que percebamos a urgência em registrar essas Línguas de Sinais indígenas da atualidade, evitando assim o seu desaparecimento da história humana.

Sobre o *corpus* em questão, QUADROS (2016a) *apud* Royer (2019, p. 81) afirma que:

Temos vários conjuntos de produções da Libras organizados a partir de projetos de pesquisa com diferentes propostas, mas todos têm em comum o registro de interações em Libras por meio de filmagens em vídeo. Alguns deles possuem também transcrições e traduções.

Tendo isso em vista, parece clara a importância de um trabalho bem sistematizado e com organização prévia para a elaboração de um *Corpus*, cuja finalidade é justamente o desenvolvimento de pesquisas em diversas áreas do conhecimento, especialmente em Linguística (teórico-descritiva ou aplicada), além de ser uma estratégia de salvaguarda da língua, no caso em tela, da Libras.

Embora exista à disposição de pesquisadores muitos projetos de pesquisa voltados para a Linguística de *Corpus*, sabemos que ainda é uma área de pesquisa com um desenvolvimento incipiente quando se trata de Língua de Sinais, especialmente no Brasil. Do que já se tem, no entanto, é importante ressaltar que os desafios referentes a todo o processo que envolve a elaboração de um *corpus* devidamente consistente é um desafio para pesquisadores da área de linguística e mesmo de tecnologia. O processo de anotação e transcrição de dados é especialmente árduo, exaustivo, por um lado, e, por outro, desafiador para os pesquisadores, sobretudo considerando-se as limitações que anotações por meio do uso de glosas impactam/limitam o trabalho de transcrição.

É pertinente destacar que a Libras é a língua de sinais majoritária usada pela comunidade surda brasileira, mas não é exclusiva. Em outros termos, hoje já se sabe que diversas outras Línguas de Sinais são usadas em comunidades onde há surdos, especialmente em aldeias indígenas e comunidades mais distantes dos grandes centros urbanos em que o acesso às associações de surdos e aos meios tecnológicos são limitados. Tanto quanto a Libras essas outras línguas deveriam ser documentadas.

Um aspecto relevante da documentação de uma língua é a publicização dos dados. Tornar os *corpora* públicos é extremamente importante em termos científicos, culturais e educacionais. Ademais, isso revela a valorização e o reconhecimento não só dos nativos/usuários dessa língua.

É a documentação que permitirá não apenas as comunidades usuárias dessas línguas, mas a toda população do país, reconhecer o valor e a riqueza de suas particularidades linguísticas e das perspectivas culturais a elas imbuídas. (QUADROS; LEITE, 2014, p. 25)

Vivemos em uma sociedade em que as pessoas ainda carecem de informações sobre as Línguas de Sinais existentes no país e mesmo acerca da Libras. Dessa forma, à medida que novas pesquisas forem se desenvolvendo e mais conhecimento for propagado sobre as Línguas de Sinais, menos preconceito e mais acesso à sociedade. Assim, os projetos de elaboração de *corpora* atendem a necessidades linguísticas, mas também de ordem sociopolítica. E mais uma vez torna-se clara e urgente a necessidade de se catalogar e registrar essas línguas em questão, pois com uma dimensão territorial tão grande, com tantas variações de uma mesma língua e nuances culturais atreladas a essas línguas.

Conforme Leite e Quadros (2014, p. 12),

é curioso notar, nesta reflexão sobre o estatuto de risco das Línguas de Sinais faladas no Brasil, como a dinâmica da dominação lingüística é sutil e perigosa. Por um lado, no âmbito dos grandes centros urbanos, a Língua de Sinais nacional se vê constantemente subjugada à ideologia de que “brasileiros falam língua portuguesa”, sendo continuamente relegada a um segundo plano dentro da política hegemônica de “inclusão educacional”, que continua enxergando essa língua como um método ou apoio para a aquisição do português e outros conteúdos curriculares ao invés de enxergá-la como um patrimônio cultural em si, a ser fomentado por seu valor intrínseco, e não apenas instrumental. Por outro lado, o prestígio crescente da Libras como “Língua de Sinais dos surdos brasileiros” acaba se tornando justificativa para que essa língua passe a ser ensinada a pessoas surdas usuárias de outras variedades de Línguas de Sinais, sem que se reconheça o valor intrínseco das Línguas de Sinais nativas, como é o caso da “cena” falada em Jaicós.

As imposições linguísticas que naturalmente existem na sociedade contra grupos minoritários é algo sério. Sem um registro linguístico, essas línguas correm o risco de desaparecer da história humana. E em concordância com Nonaka (2004), Leite e Quadros, (2014) todas as Línguas de Sinais precisam ser registradas. A Libras, embora seja a única Língua de Sinais brasileira reconhecida por lei (10.436/2002), não pode ser instrumento de sobreposição a Línguas de Sinais minoritárias. Várias comunidades no país utilizam outras Línguas de Sinais para se comunicarem e essas são tão valiosas quanto as faladas pelos usuários da Libras.

Ou seja, os processos culturais e linguísticos de um povo são genuínos e não têm status maior ou menor, melhor ou pior. Sobre o que toca à cultura surda, destaque-se o seguinte:

Cultura surda é o jeito de o sujeito surdo entender o mundo e de modificá-lo a fim de se torná-lo acessível e habitável ajustando-os com as suas percepções visuais, que contribuem para a definição das identidades surdas e das ‘almas’ das comunidades surdas. Isto significa que abrange a língua, as idéias, as crenças, os costumes e os hábitos do povo surdo. (STROBEL, 2008, p.22)

Como se pode observar, várias são as possibilidades de abordagens sobre a documentação de uma língua e os impactos disso decorrentes. A diversidade linguística do

país requer que políticas linguísticas sejam traçadas a fim de assegurar o registro das línguas, sobretudo as minoritárias, incluindo-se aí as indígenas, as quilombolas, as de imigração e as de sinais a fim de que esse patrimônio imaterial seja resguardado e preservado. Para além disso, em alguns casos, essas políticas têm que se voltar para a estratégias de revitalização dessas línguas.

### 2.3.2 A Documentação da Libras como Política Linguística

No cenário brasileiro, importantes políticas linguísticas têm se destacado em favor da Libras, especialmente com uma série de instrumentos legais que se estabeleceram como política de Estado. Aqui faremos um recorte a partir dos anos 2000. Neste intervalo de tempo, destacam-se:

- a) A aprovação da Lei 10.436, de 2002, a chamada Lei de Libras;
- b) O Decreto 5.626, de 2005, o qual regulamenta a Lei de Libras;
- c) O surgimento do Curso de Letras-Libras, em 2006, na UFSC, primeiro do País;
- d) A obrigatoriedade da disciplina de Libras nos cursos de formação de professores e de fonoaudiólogos, em conformidade com o Decreto 5.626/2005, a partir de 2006;
- e) A aplicação do ProLibras, de 2006, exame nacional com vistas à certificação para o uso e ensino da Libras, bem como para a tradução e interpretação envolvendo o par linguístico Libras/Português;
- f) A regulamentação da profissão de Tradutor Intérprete de Língua Brasileira de Sinais - Libras, Lei no 12.319, de 2010;
- g) A criação da Lei 7.387, de 2010, que institui o Inventário Nacional da Diversidade Linguística (INDL), o qual prevê a documentação das línguas sinais brasileiras;
- h) A criação de 27 cursos de Letras-Libras no País, por meio do Programa *Viver sem Limites*, de 2011, Decreto 7.612;
- i) A Lei 13.005, de 2014, que institui o PNE (Plano Nacional de Educação), principalmente a Meta 4;
- j) A formação de professores de português como segunda língua para surdos que começou a ser oferecida em 2015 pela Universidade de Brasília (UnB);

- k) A criação da Lei 13.146, de 2015, a *Lei Brasileira de Inclusão*;
- l) Exame Nacional do Ensino Médio (ENEM) em Libras 2017 a 2021;
- m) Decreto 9.508/2018 vídeo-prova em Libras para concursos públicos;
- n) Lei Nº 14.191/2021, Lei de Diretrizes e Bases da Educação Nacional), para dispor sobre a modalidade de educação bilíngue de surdos.

Mediante essas políticas públicas implementadas no Brasil, a Libras ganha cada vez mais espaço na sociedade, o que tem como implicação a necessidade de melhor conhecê-la. Os espaços acadêmicos e educacionais criam, embora sem que esse seja seu foco, registros, ou *corpora*, a todo tempo: veja, por exemplo, o caso do ENEM em Libras, do ProLibras, das diversas produções para fins de avaliação nos Cursos de Letras-Libras; além destes contextos indiretos, a Libras, a partir de 2014, passa a contar com uma documentação específica, para fins científicos, sobretudo linguísticos, por meio do Projeto Inventário de Libras, na UFSC. Este projeto tem por objetivo documentar Libras em todas as capitais brasileiras. Atualmente foi concluída a coleta de dados na UFSC, UFAL e UFC, sendo que os dados das duas primeiras instituições já estão disponíveis on-line em <https://corpuslibras.ufsc.br/>. Na UFT os dados começaram a ser coletados, mas a coleta foi interrompida em função da pandemia do COVID-19. O Instituto Nacional de Educação de Surdos - INES (Rio de Janeiro), UFAC, UFCG e Museu da Inclusão (São Paulo) são instituições que já participaram do treinamento para dar início à coleta de dados dos estados nos respectivos estados.

Aos poucos, essa importante tarefa de documentar a Libras no país vai ganhando espaço político e linguístico, de modo a se constituir um política nacional de documentação da língua, assegurando à Libras status sociolinguístico, histórico e cultural. Trata-se de um política linguística nos moldes do que propõe o INDL.

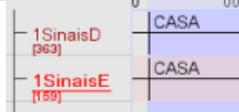
### 2.3.3 A transcrição dos dados

Com o objetivo de padronizar as anotações dos dados, foi criado o *Manual de Transcrição do Corpus da Libras do Inventário Nacional*, o qual vem sendo atualizado para deixar as anotações mais simples e consistentes. A última versão do referido Manual é 2019 e foi concebida a partir das experiências dos pesquisadores do Projeto no processo de validação dos dados pelas seguintes instituições: UFSC, UFAL, UFC, UFT e o INES. O processo de validação requer que pelo menos 70% dos dados transcritos por diferentes

transcritores apresentem consistência, a partir das normas/convenções de transcrição apresentadas no Manual de transcrição.

**Figura 5** - Tela de Manual de transcrição do *Corpus* Libras Inventário Nacional

**QUADRO SÍNTESE DO MANUAL DE TRANSCRIÇÃO DO INVENTÁRIO DE LIBRAS – Versão 12/11/2019.**

Item	Convenção	Exemplo
Glosas na Língua de Sinais	Letras maiúsculas; glosas com mais de uma palavra devem ser ligadas com hífen	COELHO NÃO-TER
Sinais D Sinais E	Trilhas de anotação dos sinais, Sinais D para os sinais que utilizam a mão direita e Sinais E para os sinais que utilizam a mão esquerda. Quando o sinal for realizado com as duas mãos, colocar o ID nas duas trilhas compartilhando o mesmo tempo.	
Apontação para pessoas	Usar o IX para apontação de pessoas.	IX
Apontação para objetos	Usar o IX para apontação de objetos.	IX
Apontação para lugares	Usar IX para a apontação de lugares.	IX
Incorporação de numeral na apontação	Pode ser usado de forma inclusiva (nós-dois, nós-três, nós-quatro) ou de forma exclusiva (você-dois, você-três, você-quatro). Colocar IX mais o número ao lado (Ex: IX2, IX3 e IX4).	IX2 IX3 IX4
Demonstrativos	Usar DEM para indicação de demonstrativos, independente do tipo.	DEM
Possessivos	Usar POSS para indicação de possessivos, independente da pessoa gramatical.	POSS
Verbos indicativos	Nomear com uma glosa ID para cada sinal; não adicionar informação sobre os referentes	DAR IR
Verbos descritivos (classificadores)	Usar apenas a glosa 'DV' sem descrição entre parêntese do evento.	DV
Palavras soletradas	Usar a glosa 'FS' seguida da palavra sem	FS(nokia)

Fonte: Proposta de Manual de transcrição do *Corpus* Libras Inventário Nacional (versão 2019)

	hifenização ou da letra entre parênteses	FS(a)
Pausa dentro da sentença	Representar a pausa da sinalização (#)	IX(si) ESCOLHER # AZUL
Gestos	Glosa com & seguido da descrição da ação entre parênteses.	&(face-braba)
Sinal não muito claro (mas que o transcritor identifica o seu significado)	Adicional [?] no final da glosa	QUERER MAÇA[?]
Sinal não muito claro (o transcritor oferece uma glosa alternativa, pois o sinal também pode ser outro)	Digitar a primeira opção de glosa, seguido por [=?ALTERNATIVA]	QUERER MAÇA[=?BOLACHA]
Sinal que não é reconhecido.	Cada sinal não reconhecido no enunciado recebe a glosa XXX (pode haver mais de um).	QUERER XXX POR-FAVOR

Fonte: Proposta de Manual de transcrição do *Corpus* Libras Inventário Nacional (versão 2019)

Vale ressaltar que mesmo com o uso do Manual de transcrição do *Corpus* da Libras, ainda existem dificuldades em padronizar as glosas no momento de transcrição, pois nem sempre há correspondência semântica entre sinal e palavra. Com isso não se quer dizer que as línguas de sinais são limitadas ou pobres, elas simplesmente diferem das línguas orais em muitos casos ou, por serem línguas relativamente novas, nem sempre dispõem de um repertório vocabular equiparado ao das línguas orais. No entanto, essa é uma questão que impacta no momento das transcrições se se toma o sistema de glosas como estratégia de transcrição. Nossa pesquisa visa contribuir para que estas dificuldades sejam discutidas, resolvidas ou amenizadas, com o intuito de tornar o trabalho de transcrição mais produtivo e menos enfadonho.

A fim de dirimir problemas nas transcrições dos dados, alguns recursos tecnológicos são usados, a exemplo do banco de sinais, plataforma que armazena sinais com suas respectivas informações gramaticais, ferramenta de busca por palavra ou por sinal. O *Signbank* pode ser facilmente acessado por qualquer pessoa que tenha acesso à Internet, pelo site <https://signbank.libras.ufsc.br>.

#### 2.3.4 A construção do *Corpus* da Libras da UFAL

O *Corpus de Libras da Grande Maceió* tem por finalidade documentar a Libras e aprimorar os recursos e estratégias para o registro desta língua, com foco no processo de documentação. A documentação de uma língua, em certa medida, já é parte da análise

linguística, uma vez que o pesquisador necessita tomar decisões acerca das transcrições, das anotações (Transcrições com uso do *Software ELAN*) as quais são baseadas em critérios necessariamente linguísticos. Os desafios para documentar uma língua por meio de *corpora* implicam o uso consistente de recursos tecnológicos a fim de que se possa lançar mão de estratégias de anotação dos dados de modo relevante, mas também de busca eficaz de dados. Nesse sentido, os aspectos tecnológicos e linguísticos estão em interface, de forma indissociável.

Como os dados que foram usados para a nossa pesquisa fazem parte de um projeto maior, sentimos a necessidade de descrever como se deu a constituição do projeto *Corpus* de Libras da Grande Maceió, empreendimento do qual fizemos parte como bolsista de iniciação científica por quatro anos.

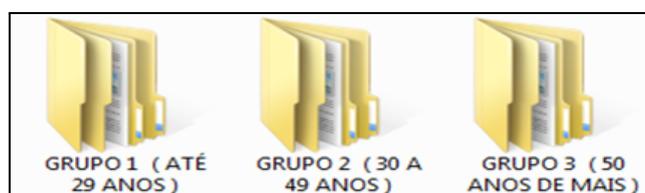
#### 2.3.4.1 Etapas para a elaboração do *corpus*

##### 2.3.4.1.1 Os sujeitos da Pesquisa

Os dados foram coletados com a participação de trinta e seis pessoas surdas, todas maiores de idade, até 29 anos (grupo 1), de 30 a 49 anos (grupo 2) e a partir dos 50 anos de idade (grupo 3), sendo homens e mulheres, independente de escolarização, desde que tivessem fluência em Libras, ou seja, pelo menos 10 anos de contato constante com a língua, e para a efetiva coleta de dados, foi necessário que assinassem o Termo de Consentimento Livre e Esclarecido - TCLE, onde existiam 12 (doze) itens com esclarecimentos, de modo que os participantes pudessem ter plena consciência do que estavam fazendo com vistas à autorização da sua imagem por meio de vídeos. Os dados foram coletados de forma sistemática, tentando-se, em que pese o uso dos diversos equipamentos necessários à captação de imagem, torná-los o mais natural possível.

Os dados dos *Corpus* da Libras da UFAL serão extraídos para análise, e estão dispostos sistematicamente como três grupos etários, distribuídos da seguinte forma:

**Figura 6** - Como três grupos



Fonte: *Corpus* de Libras: versão grande Maceió.

Dentro de cada grupo etário, há uma subdivisão sexo, como descrita a seguir:

**Tabela 1** (Mulheres surdas): Grupo 1: até 29 anos (jovens), Grupo 2: de 30 a 49 anos (adultos) e Grupo 3: mais de 50 anos (Idosos).

<b>Mulheres (18 surdas)</b>	
<b>Grupo 1</b>	Até 29 anos
<b>Grupo 2</b>	30 a 49 anos
<b>Grupo 3</b>	Mais de 50 anos

**Tabela 2** (Homens surdos): Grupo 1 até 29 anos (jovens), Grupo 2: de 30 a 49 anos (adultos) e Grupo 3: mais de 50 anos (Idosos).

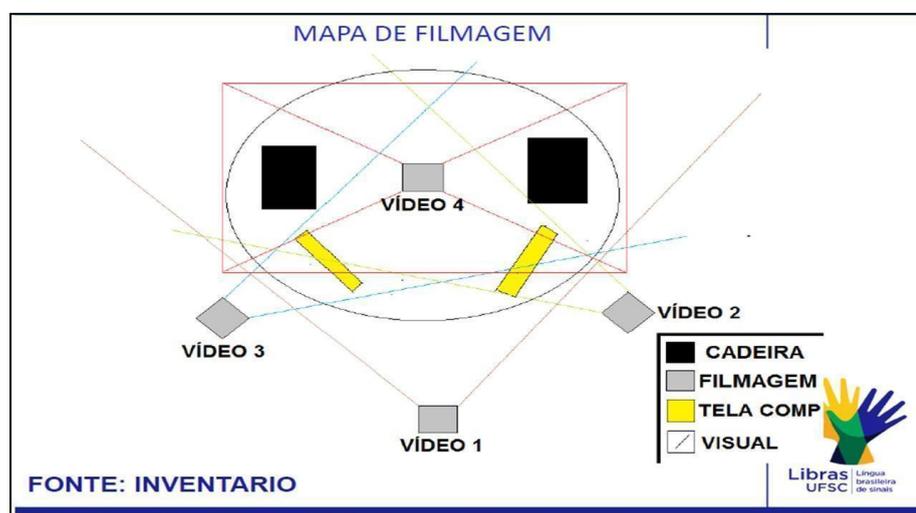
<b>Homens (18 surdos)</b>	
<b>Grupo 1</b>	Até 29 anos
<b>Grupo 2</b>	30 a 49 anos
<b>Grupo 3</b>	Mais de 50 anos

Os dados que serão analisados são os referentes aos seguintes gêneros: narrativas (04): pêra, Mr. Bean, Chaplin e Tom e Jerry. Histórias (02): frog e piteco. Conversação livre sobre temas diversos (06): trânsito em Maceió, crise política no Brasil, escolas surdos e escolas ouvintes (inclusão), tecnologia e implante coclear, associações surdos e conversa livre). Vocabulário (09): com temas; cores, calendário, família, frutas, verduras, alimentação, bebidas, profissões e regiões). Seguindo a metodologia dos dados do *Corpus* da Libras da UFAL, cabe salientar que todos os informantes são oriundos da grande Maceió ou aqui residem há pelo menos 10 anos.

### 2.3.4.2 A Coleta dos Dados

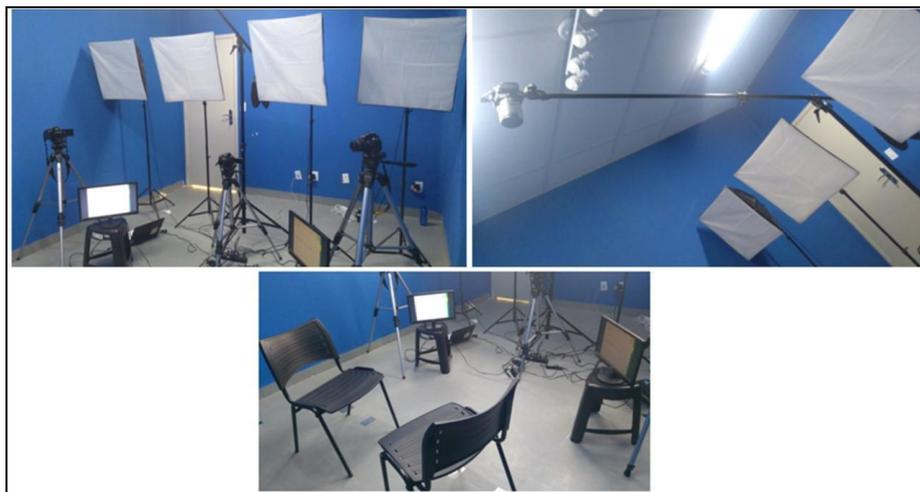
Os dados do *Corpus* de Libras da UFAL tiveram sua coleta por meio de diversos estímulos para a produção dos vídeos, a saber: reconto de narrativas, história de vida, conversação livre, vocabulário, e foi realizada de março de 2016 a agosto de 2018, com duração média de 20 a 30 minutos. As coletas se deram em um estúdio com fundo azul, iluminação fria e branca, e foram utilizadas quatro câmeras de alta definição para captação das imagens em vídeo, sendo uma câmera direcionada a cada participante, uma captando os dois informantes de frente, e outra de cima para baixo.

**Figura 7 - Espaço filmagem**



Fonte: Inventário de Libras em UFSC

Antes das gravações, o espaço é organizado de modo que se viabilize a execução dos registros, evitando-se que os participantes não tenham empecilhos durante a captação.

**Figura 8** - Organização do estúdio

Fonte: *Corpus* de Libras: versão grande Maceió.

Os participantes surdos, após assistirem a um vídeo com o TCLE em Libras, estando de acordo com o termo, fazem a assinatura do documento. Após esse importante procedimento ético, em duplas, um em frente ao outro, os sinalizantes têm acesso a um monitor, onde são apresentados os estímulos (vídeos e imagens) para as produções linguísticas, as quais são captadas por quatro câmeras em alta definição. Feito isso, os dados são descarregados em dispositivos físicos e em nuvem para posterior edição e transcrição.

**Figura 9** - Tomadas de imagens em vídeo

Fonte: *Corpus* de Libras: versão grande Maceió.

Os participantes surdos desta pesquisa, durante a coleta de dados, foram submetidos a diferentes tipos de estímulos para produção linguística: entrevistas, conversação livre, narrativas, histórias e lista de vocabulário.

Estes assuntos e temas foram padronizados pelo *Corpus* de Libras da UFSC, porém, os temas de conversação livre foram alterados com o tempo, sendo de acordo com os acontecimentos do momento, dando continuidade e atualização aos trabalhos do *Corpus* de Libras na UFAL.

#### 2.3.4.3 Seleção dos dados

Um colaborador da pesquisa, geralmente bolsista de iniciação científica ou aluno voluntário, realiza a entrevista com os dois convidados (um por vez), onde são feitas várias perguntas, como por exemplo: Nome, idade, onde nasceu, se a família sabe Libras, entre outras, de modo que a sinalização das respostas seja espontânea. Feito isso, outros temas são conversados visto que o informante já se sente mais confortável com o fato de estar em observação, permitindo uma sinalização mais natural e relaxada. Dessa forma, podemos perceber a língua sendo produzida naturalmente, sem fingimento ou controle da sinalização.

**Figura 10** - Tomadas de imagens em vídeo da entrevista



Fonte: *Corpus* de Libras: versão grande Maceió.

Nas entrevistas, cada entrevistado senta-se na frente do voluntário, responsável pela entrevista. E com o uso de um monitor, algumas imagens de cores, frutas, bebidas, entre outras, são mostradas para que, em seguida, o participante faça o sinal correspondente. Dessa forma, é registrado por meio de vídeo as variações de um mesmo sinal, entre jovens,

adultos e idosos. Na sequência, essas produções são transcritas no *Software ELAN* e então podemos ver quantas e quais foram as variações são usadas.

**Figura 11** - Tomadas de imagens em vídeo sobre vocabulário



Fonte: *Corpus* de Libras: versão grande Maceió.

Nessa etapa, dois convidados são posicionados frente ao entrevistador, um por vez, de modo que é apresentada uma imagem no monitor para o convidado dizer o sinal correspondente ao que foi visto. Enquanto o convidado que aguarda a vez, não sabe o que está sendo mostrado.

**Figura 12** - Tomadas de imagens em vídeo sobre vocabulário



Fonte: *Corpus* de Libras: versão grande Maceió.

As imagens mostradas são dos mais variados tipos, de modo que a partir do momento que o sinalizante olhar, poderá representá-las por meio dos sinais em Libras que conhece.

**Figura 13** - Tomadas de imagens em vídeo sobre vocabulário



Fonte: *Corpus* de Libras: versão grande Maceió.

Para registro das narrativas, cada participante assiste a uma narrativa diferente e, em seguida, produzem uma paráfrase do que foi visto. Na sequência, é feita a transcrição no *ELAN*, tendo como desafio fazer as anotações dos sinais e dos detalhes que são usados na produção dessa paráfrase (expressões corporais e faciais, classificadores), tendo como guia a proposta do Manual de transcrição do *Corpus* da Libras do Inventário Nacional e *SignBank*. Nesse momento, começam os desafios para os anotadores, uma vez que a produção linguística de narrativas em Libras é bastante sofisticada em termos estruturais e nem sempre a anotação por meio de glosas dá conta dessa riqueza da língua, motivo por que discutiremos mais adiante as limitações das transcrições por glosas.

**Figura 14** - Tomadas de imagens em vídeo sobre narrativa



Fonte: *Corpus* de Libras: versão grande Maceió.

Os vídeos mostrados são ricos em detalhes, propiciando o uso dos classificadores e da criatividade linguística na narração. Os seguintes estímulos foram apresentados para a produção linguística:

- O participante 1 assiste a um vídeo da Pêra (Narrativa)
- O participante 2 assistir a um vídeo do Charles Chaplin (Narrativa)
- O participante 1 assistir a um vídeo do Mr. Bean (Narrativa)
- O participante 2 assistir a um vídeo dos Tom e Jerry (Narrativa)
- O participante 1 assistir a um vídeo do FROG (História)
- O participante 2 assistir a um vídeo PITECO (Historia).

**Figura 15** - Tomadas de imagens em vídeo sobre narrativa



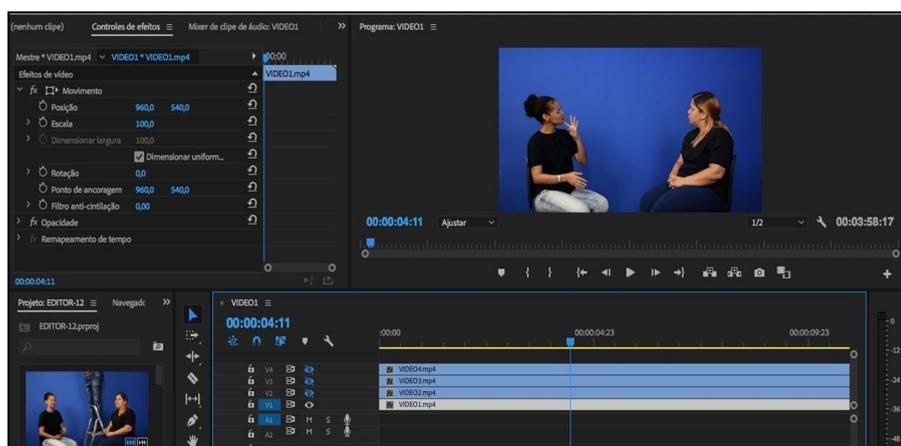
Fonte: *Corpus de Libras: versão grande Maceió.*

#### 2.3.4.4 O Tratamento dos Dados

##### 2.3.4.4.1 A Edição

Para a edição dos vídeos, foi usado o *software Adobe Premiere*, por meio do qual foram feitos todos os cortes e ajustes nos vídeos para a posterior anotação e transcrição.

**Figura 16** - Tela do *Adobe Premiere* com edição de vídeo



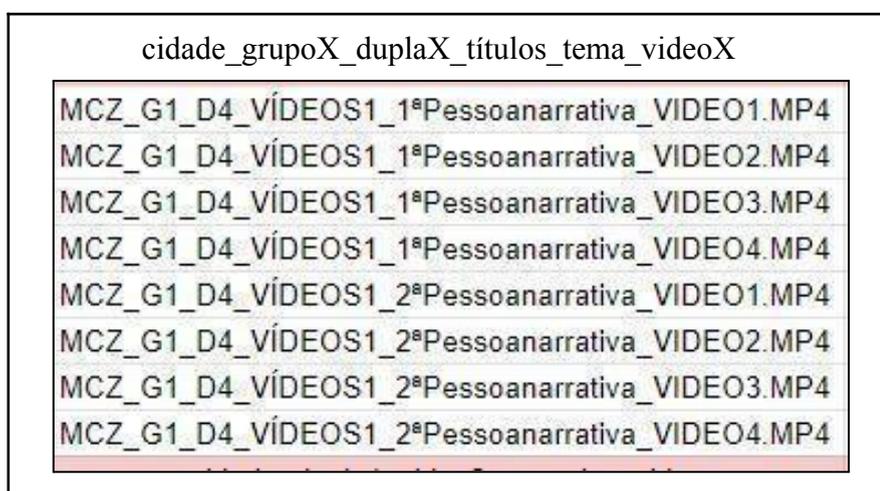
Fonte: *Corpus de Libras*: versão grande Maceió.

#### 2.3.4.4.2 A Anotação/Transcrição dos dados

Após tratamento dos dados por meio da edição, eles passam pelo processo de anotação. Os vídeos, num total de quatro (referentes às quatro câmeras de captação), agora são colocados no *software EUDICO Linguistic Annotator - ELAN*, onde é feita a transcrição por meio de glosas. O *ELAN* é um programa gratuito para anotação de arquivos de áudio e vídeo, desenvolvido pelo Instituto Max Planck de Psicolinguística em Nijmegen, Holanda.

Os dados coletados são dispostos em planilhas no *Google Drive*, onde existe a organização das gravações entre grupos e duplas de participantes. Por exemplo:

**Figura 17** - Codificação\Nomenclatura dos arquivos de vídeos transcritos

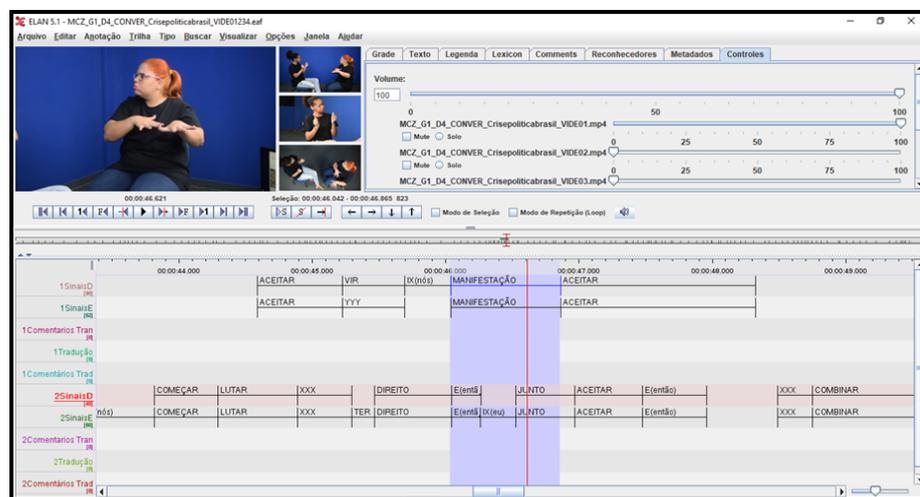


Fonte: *Corpus de Libras*: versão grande Maceió.

Nesse caso, os dados são assim codificados: MCZ (dados de Maceió), G1 (Grupo 1), D4 (Dupla 4), VIDEO1 (primeira câmera), 1ªPessoanarrativa (Informante 1, gênero narrativo), VIDEO1.MP4 (Video 1, salvo em formato MP4). Essa organização é extremamente importante para a sistematização dos dados em si, e, posteriormente, para a publicização dos dados na página do Portal da UFSC, em que todas as tags devem ser preenchidas a fim de facilitar a busca futuramente, além do registro de metadados.

Para fins de transcrição/anotação dos dados, foi criada uma estrutura no *ELAN* composta por 5 trilhas, sendo uma para cada informante, uma tradução, uma para comentários do tradutor e outra para comentários do anotador, assim dispostas: SinaisD (sinais realizados com a mão direita), SinaisE (sinais realizados com a mão esquerda), Comentários do transcritor (trilha reservada para o transcritor colocar comentários pertinentes à anotação), Tradução PB (tradução da Libras para o Português do Brasil) e, por fim, Comentários do tradutor (espaço reservado para o tradutor colocar as observações que julgar necessárias), conforme ilustrada na figura 19 a seguir e na Tabela 3.

**Figura 18** - Tela de transcrição com uso do *ELAN*



Fonte: *Corpus* de Libras: versão grande Maceió

**Tabela 3** - Dados a serem transcritos no *ELAN*

Sinalizante 1	Sinalizante 2
1 Sinais D	2 Sinais D
1 Sinais E	2 Sinais E
1 Comentários do transcritor	2 Comentários do transcritor
1 Tradução PB	2 Tradução PB
1 Comentários do tradutor	2 Comentários do tradutor

Fonte: O autor (Adaptado do Manual de Transcrição do *Corpus*)

Importante recurso para o momento de transcrição/anotação dos dados é o *SignBank* que é tomado como referência. No início do projeto, o *Identificador de Sinais* era o banco de dados tomado como base. Essa plataforma, no entanto, foi transferida e tem sido aperfeiçoada numa nova base ou banco de sinais, o *SignBank*, cuja dinamicidade de busca e possibilidade de ampliação/atualização são mais robustas do que o *Identificador de Sinais*.

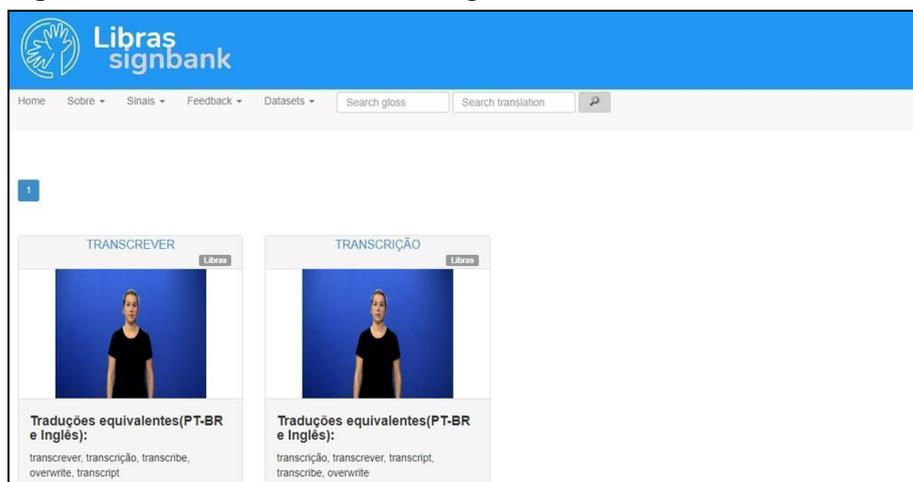
**Figura 19** - Plataforma de Identificação de Sinais

Fonte: Identificador de Sinais

O *SignBank* tem importante papel no momento das transcrições, pois é nele que os transcritores buscam as informações e referências para os sinais no momento da transcrição. Por outro lado, quando são encontrados sinais que ainda não constam no *SignBank*, os pesquisadores podem gravar e alimentar o sistema, de modo que a atualização é constante, tornando-o mais amplo e consistente. Tanto sinais novos que ainda não constam no *SignBank*, quanto sinais que apresentam variação são inseridos e isso facilita as novas

transcrições, pois trabalhos futuros já terão acesso a um *SignBank* mais rico em termos de quantidade e qualidade de dados. O site *Libras SignBank* é o seguinte: <https://signbank.libras.ufsc.br/>.

**Figura 20** - Plataforma de Libras *SignBank*



Fonte: *Libras SignBank*

A organização prévia e posterior para o tratamento dos dados é muito importante. Fazemos uso da planilha do *Google Drive*, cuja disposição é padronizada nos *Corpora* de Libras no Brasil, onde são colocados dados de identificação do transcritor, data de transcrição e o nome do arquivo. De modo que exista uma organização e facilitação na busca dos arquivos.

**Figura 21 - Registro dos dados da transcrição em caderno**

GRUPO ____ (_____)					
DUPLA ____ (_____)					
Participante - 1 (_____)	Transcritor	Pronto	Tradutor	Pronto	Revisão
1. Entrevista					
2. Primeira narrativa PERA					
3. Segunda narrativa MR. BEAN					
4. História FROG					
5. Vocabulário					
Participante - 2 (_____)	Transcritor	Pronto	Tradutor	Pronto	Revisão
1. Entrevista					
2. Primeira narrativa CHARPIN					
3. Segunda narrativa TOM e JERRY					
4. História PITECO					
5. Vocabulário					
Conversação (_____)	Transcritor	Pronto	Tradutor	Pronto	Revisão
1. Trânsito em Maceió					
2. Crise política no Brasil					
3. Escola surdos X Escola ouvintes					
4. Tecnologia X Implante coclear					
5. Associações Surdos					
6. Conversa livre					

Fonte: *Corpus* de Libras: versão grande Maceió

Para fazer uma transcrição em *ELAN*, primeiramente, é registrado o nome do transcritor, data de transcrição (pronto), nome do tradutor, data de tradução (pronta) e da revisão em um caderno (veja a figura 22 acima). Posteriormente, dispor destas informações numa planilha no *Google Drive*, esta tabela está organizada de acordo com a faixa etária de cada grupo, seguindo o exemplo da planilha abaixo (figura 23). Dessa forma, o registro é mais acelerado e mais seguro de ser realizado. Os nomes reais dos participantes são agregados respectivamente por: Grupo 1 - até 29 anos, Grupo 2 - 30 até 49 anos e Grupo 3 - 50 anos de mais. O grupo 1, 2 e 3 masculino (dupla de 1 a 3) e o grupo 1, 2 e 3 feminino (dupla de 4 a 6).

Figura 22 - Organização pelos vídeos (planilhas no *Google Drive*)

GRUPO 1 (até 29 anos)	DUPLA 1	DUPLA 2	DUPLA 3	DUPLA 4	DUPLA 5	DUPLA 6
grupo11dupla1	grupo11dupla2	grupo11dupla3	grupo11dupla4	grupo11dupla5	grupo11dupla6	
grupo21dupla1	grupo21dupla2	grupo21dupla3	grupo21dupla4	grupo21dupla5	grupo21dupla6	
grupo31dupla1	grupo31dupla2	grupo31dupla3	grupo31dupla4	grupo31dupla5	grupo31dupla6	

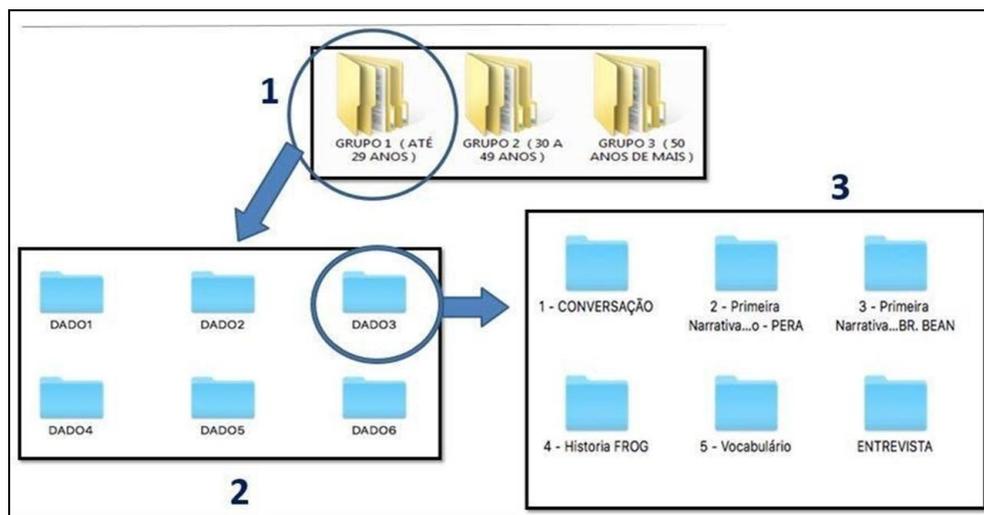
GRUPO 1 - DUPLA 4	NOME PARTICIPANTE	VIDEOS	TRANSCRITOR	EAF pronto	TRADUTOR	Tradução Pronto	REVISÃO final
		cidade_dupla4_entrevista_video	TRANSCRIÇÃO			TRADUÇÃO (DATA)	REVISÃO
	MAIARA E CLÁUDIA	MCZ_G1_D4_1entrevista_VIDEO1.MP4	Ewerton e May				
	MAIARA	MCZ_G1_D4_1entrevista_VIDEO2.MP4	Ewerton e May				
	CLÁUDIA	MCZ_G1_D4_1entrevista_VIDEO3.MP4	Ewerton e May				
	MAIARA E CLÁUDIA	MCZ_G1_D4_1entrevista_VIDEO4.MP4	Ewerton e May				
	MAIARA E ISABEL	MCZ_G1_D4_2entrevista_VIDEO1.MP4					
	MAIARA	MCZ_G1_D4_2entrevista_VIDEO2.MP4					
	ISABEL	MCZ_G1_D4_2entrevista_VIDEO3.MP4					
	MAIARA E ISABEL	MCZ_G1_D4_2entrevista_VIDEO4.MP4					
		cidade_dupla4_conversacao_tema_video					
	ISABEL E CLÁUDIA	MCZ_G1_D4_CONVER_Transito_VIDEO1.MP4	Ewerton	PRONTO (03/07/2018)			
	ISABEL	MCZ_G1_D4_CONVER_Transito_VIDEO2.MP4	Ewerton	PRONTO (03/07/2018)			
	CLÁUDIA	MCZ_G1_D4_CONVER_Transito_VIDEO3.MP4	Ewerton	PRONTO (03/07/2018)			
	ISABEL E CLÁUDIA	MCZ_G1_D4_CONVER_Transito_VIDEO4.MP4	Ewerton	PRONTO (03/07/2018)			
	ISABEL E CLÁUDIA	MCZ_G1_D4_CONVER_Crisepolitcabrasil_VIDEO1.MP4	Sergio	PRONTO (11/07/2018)			
	ISABEL	MCZ_G1_D4_CONVER_Crisepolitcabrasil_VIDEO2.MP4	Sergio	PRONTO (11/07/2018)			
	CLÁUDIA	MCZ_G1_D4_CONVER_Crisepolitcabrasil_VIDEO3.MP4	Sergio	PRONTO (11/07/2018)			
	ISABEL E CLÁUDIA	MCZ_G1_D4_CONVER_Crisepolitcabrasil_VIDEO4.MP4	Sergio	PRONTO (11/07/2018)			
	ISABEL E CLÁUDIA	MCZ_G1_D4_CONVER_Escolasurdolouvinte_VIDEO1.MP4					
	ISABEL	MCZ_G1_D4_CONVER_Escolasurdolouvinte_VIDEO2.MP4					

Fonte: *Corpus* de Libras: versão grande Maceió

### 2.3.4.4.3 O Armazenamento dos Dados

Outra etapa extremamente importante para a organização de um *corpus* é o armazenamento de dados. Armazenar os dados requer planejamento, infraestrutura e, no caso de Línguas de Sinais, muito espaço para armazenamento dos vídeos. O procedimento é feito em dispositivos físicos e em nuvem. Posteriormente, todos os dados serão transportados e armazenados no servidor da UFSC e disponibilizados on-line na página do *Corpus* de Libras (<http://www.corpuslibras.ufsc.br>), ao término dos projetos. Além disso, serão armazenados no *The Language Archive* (<https://tla.mpi.nl>), arquivo internacional de dados linguísticos que preza por segurança na manutenção dos dados. o *The Language Archive* armazena os dados em diferentes país para manter a segurança dos dados.

**Figura 23** - Sistematização dos dados



Fonte: *Corpus* de Libras: versão grande Maceió.

No site do *Corpus* de Libras, procure o menu e clique em "Acervo" para que apareça referente ao inventário da Libras da grande Maceió. O site é: (<https://corpuslibras.ufsc.br/dados>). Veja na figura 25, abaixo:

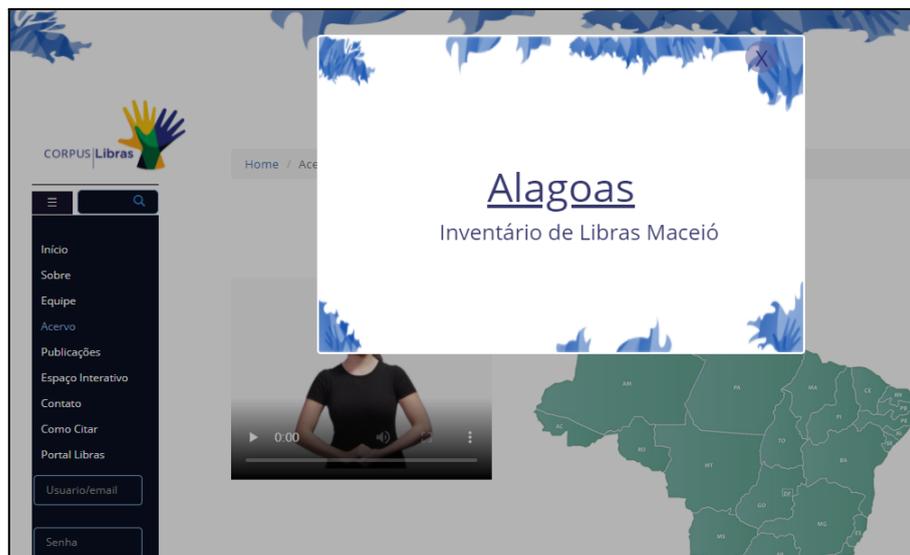
**Figura 24** - Disponibilizados on-line na página do *Corpus* de Libras



Fonte: *Corpus* de Libras - UFSC

Existe conteúdo e acesso às informações para fins de estudos e apreciação, podendo ajudar no desenvolvimento de pesquisas, onde os profissionais têm à disposição uma seção própria do inventário de Libras de Alagoas (UFAL) e de Santa Catarina (UFSC) e assim por diante. Futuramente, mais estados também poderão ter seus *corpora* catalogados no site.

**Figura 25** - Disponibilizados on-line na página do *Corpus* de Libras Grande Maceió



Fonte: *Corpus* de Libras - UFSC

A coleta de dados está organizada da seguinte forma: ao digitar o nome do formato do vídeo do arquivo, ID dados, a data de coleta, o nome do projeto, o nome do responsável, a categoria, tags e nome dos dois participantes. Depois disso, são colocados os vídeos do *corpus* da Libras e, em seguida, os vídeos são publicados na página do *Corpus*.

**Figura 26** - Organização de coleta de dados em vídeos na página do *Corpus*



Fonte: *Corpus* de Libras: versão grande Maceió

A ampliação do trabalho de transcrição/anotação dos dados, bem como a publicização desse trabalho, favorece as pesquisas na área de Linguística de *Corpus*, além de estimular outros pesquisadores a se encorajarem para a realização de elaboração de *corpora* de seus estados de origem. A ideia básica do projeto é que se possa ter um inventário nacional da Libras, portanto, com dados de todos os estados brasileiros.

Foi necessário descrever todo o percurso metodológico, da preparação à publicização do *Corpus*, uma vez que a nossa análise será pautada nos dados do *Corpus* de Libras da Grande Maceió. Mas, para além disso, toda a minha formação acadêmica se deu nesse contexto, em que por quatro anos fiz parte como bolsista de iniciação científica do projeto do *Corpus*. No próximo capítulo, dedicaremos a nossa atenção à análise de dados do *Corpus* a partir do uso de glosas, com as vantagens, mas, sobretudo, as dificuldades que o uso de glosas pode trazer para o trabalho de elaboração de *corpora* de Línguas de Sinais.

O trabalho de transcrição dos dados foi iniciado e contou com a participação dos bolsistas do projeto, no entanto, muitas dificuldades de diferentes naturezas surgiram fazendo com que o trabalho de transcrição não tenha sido concluído até o presente momento. Estas dificuldades estão relacionadas ao fato de que alguns bolsistas, como eu, terem saído do projeto, pois concluíram a graduação. Somado aos problemas metodológicos que serão melhor explicados na seção seguinte, a pandemia tornou este processo inviável, pois não foi possível prosseguir com o projeto ativamente. No momento em que estamos voltando às atividades presenciais com novos alunos no curso Letras-Libras, o projeto poderá voltar a acontecer, mas as questões procedimentais se mantêm. Desta forma, esta pesquisa irá observar estas questões e propor soluções para possíveis problemas com o intuito de contribuir com as próximas etapas, não só deste projeto, mas de tantos outros que estão surgindo no Brasil.

### **III. CAPÍTULO - Metodologia da análise das transcrições**

Nesta seção, explicaremos a coleta de dados desta pesquisa, apresentando os participantes, os instrumentos e os procedimentos metodológicos para a coleta. Com o objetivo de mapear os desafios encontrados na transcrição dos dados até o momento, nós decidimos realizar uma pesquisa qualitativa para saber como este processo aconteceu com os transcritores do *Corpus*. Foi conversado com cada um deles individualmente, e feito um encontro com dois dos 3 demais em conjunto, também pelo *ZOOM*. Infelizmente não conseguimos nos encontrar os 4 ao mesmo tempo. A partir dos relatos, anotações e das próprias transcrições realizadas, conseguiremos identificar os principais problemas encontrados pelos transcritores, e, a partir de então, sugerir caminhos para saná-los.

### 3.1 Coleta de dados

A coleta de dados foi dividida em duas etapas. Inicialmente foram feitas entrevistas semi-estruturadas com os transcritores para saber como foi este processo de transcrição, quais os procedimentos adotados e quais desafios e estratégias foram criados durante o processo de transcrição. Após a entrevista, analisamos as transcrições realizadas até o momento, juntamente com os transcritores, com o intuito de encontrar problemas de transcrição e possíveis estratégias para cada problema encontrado. Todo este processo será explicado detalhadamente nesta seção.

#### 3.1.1 Participantes da pesquisa

Como a pesquisa está restrita à construção do *Corpus* da Libras de Maceió, nós procuramos os transcritores desse *corpus*. Como eu também fui um dos transcritores, irei me colocar como participante, pois estarei, durante as entrevistas, contribuindo com a discussão e apresentando a minha experiência, desafios e estratégias enquanto transcritor. Sendo assim, somos 4 participantes da pesquisa. O quadro abaixo apresenta os participantes e suas respectivas informações. Embora bolsistas ouvintes também compusessem o quadro de bolsistas do Projeto, para efeito desta pesquisa foram eleitos 04 (quatro) participantes surdos.

**Tabela 4** - Participantes

PARTICIPANTE	idade	graduação no Curso Letras-Libras	período ativo no projeto	surdo/ouvinte
Atalaia	27	concluída	2015 - 2019	surdo
Arapiraca	26	em andamento	2018 - 2019	surdo
Maceió	29	concluída	2014 - 2019	surdo
Eu	32	concluída	2014 - 2018	surdo

Fonte: Elaborado pelo autor.

Detalhes sobre a preparação dos transcritores para a anotação dos dados, assim como o envolvimento de cada um no projeto serão apresentados nos resultados, pois serão

informações coletadas durante a pesquisa. A seguir apresentaremos o instrumento de coleta e o procedimento a ser realizado para coletar os dados.

### 3.1.2 Procedimentos e instrumentos da coleta de dados

Como já mencionado, a coleta foi realizada a partir de uma entrevista semi-estruturada. As perguntas foram feitas com o intuito de alcançar os objetivos específicos desta pesquisa. As perguntas foram feitas em Libras pelo *ZOOM*, uma vez que tanto eu quanto os demais participantes são surdos. Foi feito um questionário individual com cada participante e uma versão reduzida em grupo, pois é possível que a interação no momento da resposta possa fazê-los lembrar de problemas ou estratégias que eles não lembraram durante a entrevista individual. A versão em português das perguntas se encontra a seguir:

- 1 Quais os vídeos que você transcreveu?
- 2 Quantos minutos de transcrição?
- 3 Quantas horas mais ou menos você demorou para fazer a transcrição?
- 4 Você fez cursos para aprender a transcrever no *ELAN*?
- 5 Você usou o manual (qual versão) para transcrever o *ELAN*?
- 6 Como foi a experiência de seguir o manual quando foi fazer a glosa?
- 7 Qual foi sua experiência ao utilizar o *SignBank* - Libras?
- 8 Quais foram os desafios que você encontrou quando estava fazendo as transcrições?
- 9 Quais estratégias você desenvolveu para lidar com estes desafios?
- 10 Você contribuiu para a criação de novos tokens para o banco da Libras?

Vale lembrar que esta é uma versão em português das perguntas feitas em Libras, e que mesmo as perguntas feitas em Libras podem não ser idênticas de participante para participante, pois deixarei o participante livre para fornecer quaisquer informações, de tal forma que uma resposta pode servir para mais de uma pergunta, ou que uma resposta possa ajudar a entender mais facilmente a pergunta seguinte.

Após este momento das perguntas, foi aberto *ELAN* no *ZOOM* e os participantes me mostraram exemplos de problemas encontrados em suas transcrições e quais as estratégias de anotação para cada problema, ou que caminhos metodológicos eles desenvolveram para anotar dados mais complexos. Nesse momento, o participante esteve livre para mostrar

quaisquer exemplos de anotação, tanto de transcrição de glosa, quanto de comentários no *ELAN*, ou de detalhes de sinalização nos vídeos.

### 3.1.3 Tratamento dos dados (transcrições)

Tanto a entrevista, quanto este momento de ver as transcrições serão gravados para ajudarem no momento da análise. Os vídeos gravados no *ZOOM* foram traduzidos para o português, para facilitar a escrita da análise. Na versão em português, fiz a categorização das respostas para cada pergunta para facilitar a análise. Quanto aos exemplos apresentados no *ELAN*, foi anotado o arquivo apresentado pelo participante e o tempo de cada exemplo para conferência posteriormente. Como esta etapa também foi gravada, a versão em português deste momento contará com estas informações sobre os arquivos apresentados. A depender da quantidade de exemplos apresentados, foram quantificadas as ocorrências para mostrar resultados quantitativos sobre os tipos de problemas apresentados, pois é possível que muitos problemas de natureza distintas apareçam e não consigamos dar conta de todos os problemas. Neste caso, focaremos nos mais recorrentes.

### 3.1.4 Resultados

Será apresentado um relato de como foi realizada a coleta, pois é possível que outras informações surjam para contribuir com a compreensão deste processo de transcrição. Foram apresentados os resultados para as primeiras cinco perguntas em formato de tabela, por serem perguntas mais objetivas, enquanto que foram destacados e explicados as principais respostas para as demais perguntas, de modo a construir uma base para a análise dos dados na seção seguinte. Enfim, foi feito um apanhado dos principais problemas encontrados durante a transcrição dos dados e as possíveis estratégias.

### 3.1.5 Análise e discussão

Na análise dos dados, discutimos a natureza dos problemas encontrados para tentar categorizar os problemas em tipos de problemas, com o intuito de propor soluções mais gerais para estas questões. Com base nos tipos de problema, buscaremos encontrar soluções baseadas na literatura estudada e nas respostas dos próprios participantes.

## IV. CAPÍTULO - Resultados

Iniciaremos este capítulo apresentando um relato sobre a experiência da coleta de dados com os demais transcritores. Este relato será importante para entendermos como aplicamos a metodologia e quais possibilidades surgiram durante este percurso. Após o breve relato, seguiremos com a apresentação dos resultados da investigação, que servirão como base para a nossa discussão sobre os problemas e estratégias de transcrição desenvolvidas pelos transcritores e a partir das reflexões desta pesquisa.

### 4.1 Relato de experiência sobre a coleta de dados

O primeiro fato a ser destacado nesta seção é o pesquisador também ser participante na pesquisa. Durante todos os 4 anos de graduação estive engajado neste Projeto como bolsista do PIBIC. Inicialmente eu não fazia ideia do que era fazer pesquisa, ou coletar dados. Com o tempo, após muitas reuniões com meu orientador, viagens para formações e reuniões com os demais integrantes do projeto nacional, pude adquirir conhecimento e experiência para colocar o projeto em prática juntamente com os demais colegas. Além disso, a graduação em Letras-Libras foi fundamental para que eu desenvolvesse o conhecimento linguístico e científico para desenvolver o projeto. Relembrar das formações e do período de transcrição juntamente com os colegas me fez revisitar este momento de aprendizado e crescimento acadêmico.

Também não pude deixar de refletir sobre o quanto o processo de transcrição do *corpus* é complexo, pois tivemos diversas formações e reuniões e, ainda assim, muitas dúvidas perduram até hoje. Todas as etapas da criação do *corpus* foram executadas e concluídas, ainda que com alguns percalços, mas a etapa da transcrição trouxe questões que, juntamente com fatores, como já mencionado, como a pandemia e a conclusão de curso de alunos do projeto, estagnaram o projeto. Os resultados e discussões desta investigação contribuirão não somente para o andamento deste projeto, mas também com os demais projetos a serem desenvolvidos de coleta e transcrição de dados em Libras.

Inicialmente, minhas reflexões sobre os problemas de transcrição se limitavam à minha experiência como transcritor. Após a qualificação, com a contribuição da banca e após conversas com meu orientador, entendemos a importância de contar com a experiência dos demais transcritores. Inicialmente, pensei em fazer um encontro presencial com os

transcritores, mas isto não foi possível por conta dos horários livres que não se cruzavam. Fizemos, portanto, encontros pelo *ZOOM*.

A cada pergunta, percebia que foi uma escolha acertada incluir os demais transcritores na pesquisa, pois eles trouxeram contribuições que eu nem havia pensado. Além disso, muito do que eu já havia pensado como problema e estratégia foi ratificado pelos colegas, o que trouxe ainda mais segurança para o caminho que estamos trilhando neste trabalho. Enquanto participante da pesquisa como transcritor, pude também a cada entrevista, contribuir para a discussão trazendo as minhas indagações e estratégias para lidar com os problemas que surgiram durante a transcrição. Também percebi que a minha participação na entrevista fez com que os demais colegas se lembrassem de detalhes sobre as suas experiências que teriam passado despercebidos se eu não tivesse participado.

Por fim, durante a entrevista, uma das colegas informou que existe um caderno de anotações feito por pelos transcritores após a minha saída do projeto, com dúvidas e descrições em *SignWriting* (Escrita de Sinais) para salvaguardar a maneira como o sinal foi produzido e posteriormente o sinal ser incluído no *SignBank*. Após este breve relato, pudemos entender a importância desta metodologia para alcançar os objetivos e como os resultados a serem apresentados na próxima seção foram obtidos.

## **4.2 Entrevista com os transcritores**

Nesta seção, apresentaremos as respostas dos participantes à entrevista - como as primeiras perguntas são mais objetivas, apresentaremos as respostas das 5 primeiras perguntas - e, em seguida, relacionaremos estas respostas com as demais, pois entendemos que as primeiras respostas estão diretamente relacionadas com as últimas. Todas as respostas e comentários apresentados aqui serão discutidos no capítulo seguinte, que trata da discussão.

### **4.2.1 Resposta ao questionário**

Como as 5 primeiras perguntas são mais objetivas, apresentaremos um quadro com estas respostas para, em seguida, apresentarmos as respostas das demais perguntas dadas por cada participante, relacionando as respostas da segunda parte com as da primeira. As cinco primeiras perguntas do questionário foram:

- 1 Quais os vídeos que você transcreveu?
- 2 Quantos minutos de vídeo você transcreveu?
- 3 Quantas horas mais ou menos você demorou para fazer a transcrição?
- 4 Você fez cursos para aprender a transcrever no *ELAN*?
- 5 Você usou o manual (qual versão) para transcrever o *ELAN*?

Para estas perguntas, os participantes responderam:

**Tabela 5 - Participantes**

	<b>Pergunta 1</b> Quais os vídeos que você transcreveu?	<b>Pergunta 2</b> Quantos minutos de transcrição?	<b>Pergunta 3</b> Quantas horas mais ou menos você demorou para fazer a transcrição?	<b>Pergunta 4</b> Você fez cursos para aprender a transcrever no <i>ELAN</i> ?	<b>Pergunta 5</b> Você usou o manual (qual versão) para transcrever o <i>ELAN</i> ?
Atalia	MCZ G1 D6 VÍDEOS1 1ª Pessoanarrativa	1:43 min.	3h até 4 horas	Sergio, Miriam, Ewerton e Evely orientaram. Formação na UFSC e na UFAL.	SIM
Arapiraca	MCZ G2 D4 VÍDEOS1 1ª Pessoanarrativa	1:34 min.	2 horas	Miriam e Benício orientaram. Formação na UFAL.	SIM
Maceió	MCZ G1 D4 Entrevista 1	15:50 min.	3 dias	Sergio tutorial, Youtube. Formação na UFSC e na UFAL.	SIM

Eu	MCZ G1 D4 VÍDEOS1 1ª Pessoanarrativa	4:07 min.	4 horas	Miriam, Ronice, Ewerton e Evely orientaram. Formação na UFSC e na UFAL.	SIM
----	---	-----------	---------	--	-----

Fonte: Elaborado pelo autor. Pergunta 1 vídeo disponível em: <os/dado/porprojeto/Invent%C3%A1rio%20de%20Libras%20Macei%C3%B3>. Acesso em: 22 de junho de 2022.

A partir destas respostas, percebemos o quão lento é o processo de transcrição, pois os transcritores levaram em média uma hora para transcrever um minuto. Este tempo não é fora do padrão de tempo de transcrição do projeto de nível nacional, ou seja, os outros transcritores dos *corpora* dos outros estados também levaram essa média de tempo para transcrever os vídeos, pois possivelmente se depararam com as mesmas dificuldades. Já para Quadros (2016):

A transcrição é um processo que demanda um grande investimento de tempo e dedicação, particularmente nas pesquisas com línguas de sinais, que não possuem um sistema de escrita convencional e plenamente adaptado ao computador. Uma estimativa geral relatada em projetos de pesquisa com línguas de sinais é a de uma hora de trabalho de transcrição para cada minuto de gravação (Quadros, 2016, p. 21-22).

Algo ainda a ser considerado inicialmente são as respostas da pergunta 4. Todos os transcritores receberam suporte do projeto para realizar as transcrições, dos próprios colegas, dos professores e de formações de iniciativa do projeto nacional. Isto implica dizer que os problemas a serem discutidos nesta pesquisa não dizem respeito a um despreparo ou desatenção por parte dos organizadores do projeto, mas ao próprio processo de construção de um *corpus* que deve passar por adequações a partir de investigações como esta para que o processo de transcrição se torne mais produtivo e prático.

A quinta pergunta nos leva para uma discussão um pouco mais abrangente que envolve as demais perguntas e respostas, pois todos os transcritores utilizaram o manual no processo de transcrição, mas ainda assim questões técnicas de como transcrever particularidades da Libras perduraram até a presente pesquisa. Desta forma, cabe a nós agora entender quais são essas dificuldades e por que o manual não foi o suficiente para sanar todas as dificuldades dos transcritores.

Apresentaremos agora as respostas para as cinco últimas perguntas, dispostas novamente abaixo:

- 6 Como foi a experiência de seguir o manual quando foi fazer a glosa?
- 7 Qual foi sua experiência ao utilizar o *SignBank* - Libras?
- 8 Quais foram os desafios que você encontrou quando estava fazendo as transcrições?
- 9 Quais estratégias você desenvolveu para lidar com estes desafios?
- 10 Você contribuiu para a criação de novos tokens para o banco da Libras?

Todas as respostas foram transcritas para o português por um colaborador surdo, portanto, iremos citar as respostas em português e, se preciso, detalharemos como um ou outro transcritor utilizou sinais específicos que podem nos ajudar a compreender melhor os sentidos das respostas e a encontrar as respostas que procuramos. Para apresentar as respostas, destacaremos, a cada pergunta, partes chave da resposta de cada transcritor e faremos um pequeno resumo do que cada transcritor respondeu. No capítulo da discussão possivelmente traremos algumas citações de respostas que podem estar diretamente ligadas à discussão.

## **6 Como foi a experiência de seguir o manual quando foi fazer a glosa?**

A transcritora Atalaia falou da sua experiência já apresentando problemas de transcrição. Aqui ela apresentou dois principais problemas, o primeiro relacionado às diferenças de modalidade quando disse:

“A estrutura da Libras para traduzir para a língua portuguesa, para ficar à mesma maneira, foi um pouco difícil, [...] é dificultoso porque podem variar glosas, não só um [...] a mesma Libras, porque tem sua língua natural, por isso, temos sorte por ter o manual que pode utilizar DV para justificar sobre isso, se não tivesse o uso de manual como iria usar pelo glossário? É difícil.”

Ela exemplificou alguns sinais que podem ser glosados como “CORTAR” e “PEGAR” no português, mas que na verdade não apresentam um único sinal na Libras. O segundo problema apresentado pela transcritora foi em relação às diferenças fonológicas entre os sinais, trazendo como exemplo o sinal “DIFÍCIL”. Para isso, a transcritora propõe:

eu coloco em “XXX”, depois reúno com os pesquisadores para saber a decisão sobre validação das glosas e então pode-se modificar pelo manual para qualquer sinal, se for complicado é necessário anotar até a variação linguística, quer dizer que o sinal ficou um pouco diferente do que nós usamos [...] DIFÍCIL(intensidade): os quatros dedos abertos com movimento de zig-zag na testa (ver figura abaixo), pesquisei no *SignBank*, mas não existe esse sinal, existe o sinal comum para a palavra DIFÍCIL, que tem distinção na configuração de mão.

**Figura 27** - Sinal de DIFÍCIL(intensidade)



Fonte: Imagem disponível em: <<https://youtu.be/zmqR8-ISfOE>>. Acesso em: 31 de julho de 2022.

Fonte: Elaborado pelo autor.

A transcritora já havia apresentado esta dificuldade com a variação fonológica na primeira pergunta no começo da entrevista. Ela também pontuou que teve dificuldades para utilizar o manual por conta destes casos mais complexos.

A transcritora Arapiraca, por sua vez, apenas explicou que não contou muito com o manual para transcrever os sinais, mas com o *SignBank*. Ela disse: “eu ficava confuso para usar esse manual ou no *SignBank*, porque nós ainda usamos esse manual até hoje”. Quando perguntei se ela também marcava os sinais sem glosa com “XXX”, ela disse que lembrava que fazia desta maneira, mas não lembrava mais de quais sinais ela não sabia a glosa.

Por fim, o transcritor Maceió também afirmou utilizar mais o *SignBank*, e explicou que utilizava o código “XXX” como alternativa quando ele não via uma glosa no *SignBank* para os sinais a serem transcritos. Quando perguntei quais seriam os próximos passos após marcar os sinais desta forma, ele respondeu: “então, parei. Quando colocar tudo em ‘XXX’, depois eles vão gravar pelos sinais e enviam para o grupo que vai colocar pelo *SignBank* do

sistema”. Como ele já havia concluído a graduação<sup>4</sup>, os novos transcritores teriam o papel de rever estes sinais e gravar os sinais com as glosas no *SignBank*.

Sobre o manual especificamente, o transcritor respondeu à pergunta 5 que:

desde 2015, a manual tinha muita lista, quer dizer, este assunto sobre sinais de pontuação é cheio de regras para fazer [...] pois, os pesquisadores validaram na reunião que deram uma tomada decisão para tirar esse assunto do manual, porque esse manual tem IX é bem resumido para que possamos usar essa facilitação em apontamentos na trilha e não precisa ser usado em pessoas pronomes e nem usar parênteses não é necessário.

Ou seja, em sinais de apontação, eles não precisam mais procurar o pronome em português correspondente, mas somente utilizar o código IX. Sobre isso, o autor em outro momento também mencionou outra mudança acordada em reunião relacionada às descrições visuais (DV). No manual, quando alguma DV - também conhecida como classificador ou ação construída - aparecia no *corpus*, eles deveriam glosar com o código DV e entre parênteses explicar o que estava sendo descrito com os classificadores. Entretanto, após reunião com o projeto nacional, foi decidido que a explicação entre parênteses não seria mais necessária. Desta forma, o transcritor só deveria glosar o código DV nestes casos. Para o transcritor, ambas mudanças facilitaram o trabalho de transcrição. Ainda assim, o autor explicou que com o *SignBank*, muitas orientações do Manual não eram mais necessárias, pois o transcritor só precisava seguir o *SignBank*.

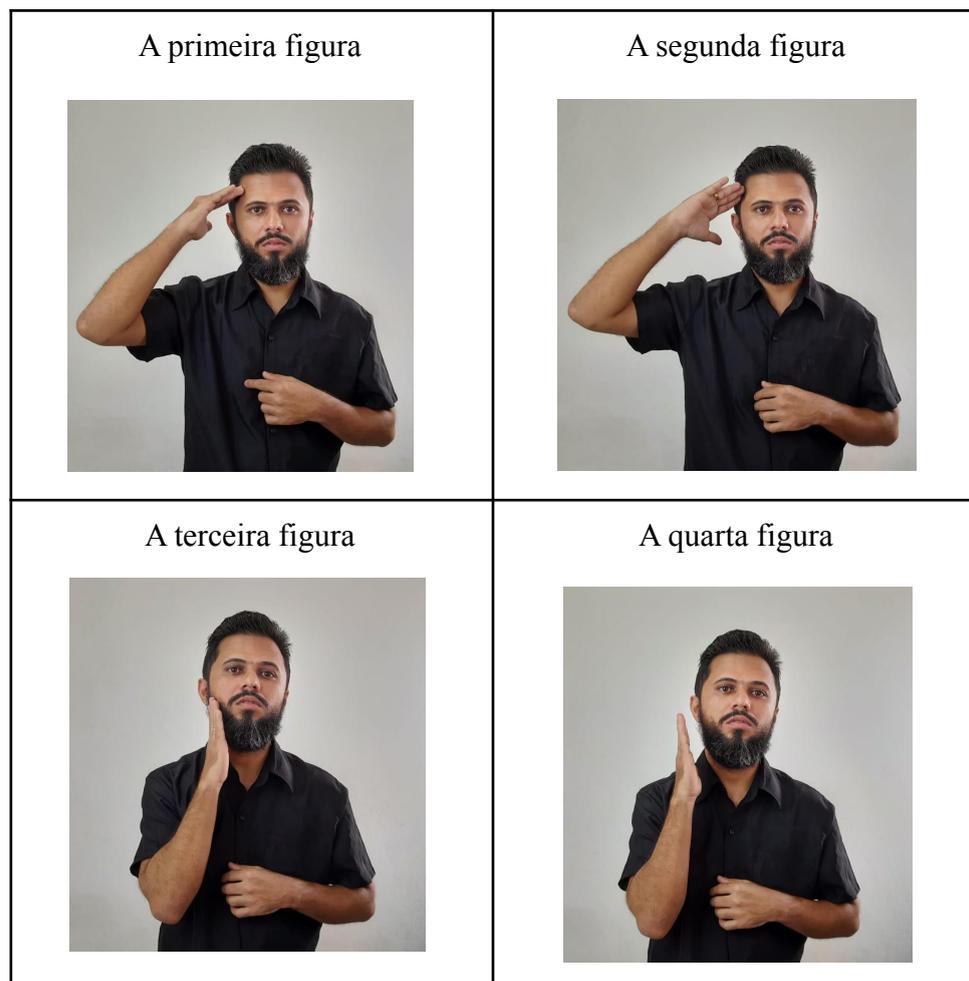
## **7 - Qual foi sua experiência ao utilizar o *SignBank* – Libras?**

A transcritora Atalaia afirmou que o *SignBank* facilitou substancialmente o trabalho de transcrição, pois procurar o sinal no *SignBank* era mais seguro do que seguir o manual e decidiu sobre qual glosa colocar. Ainda assim, a transcritora afirmou que o banco poderia ser mais flexível, quando disse

Era bom ter uma flexibilidade no *SignBank*, poderia melhorar mais ainda esta contribuição pela transcrição, para ter mais crescimento, por exemplo, um sinal para ‘ENTENDER’ segundo a primeira figura com movimento no lado da testa; e conforme a segunda figura com mão aberta para esfregar com repetição movimento no lado da testa; é possível ver evolução desse sinal, que evolui com movimento no lugar da bochecha por alofone, inclusive é possível observar no momento de produtividade de um participante durante uma gravação, de acordo com a terceira figura a mão aberta com movimento baixo na bochecha se tocando, naturalmente. Ainda tem um sinal que um informante fez sinalizado afastando da bochecha com mão aberta no segundo a quarta figura. Então achava que poderia colocar estes exemplos para ser usado as variações desses sinais no *SignBank* para que possa utilizar alofone.

---

<sup>4</sup> A equipe de trabalho do *corpus* é formada por graduandos e bolsistas de iniciação científica.

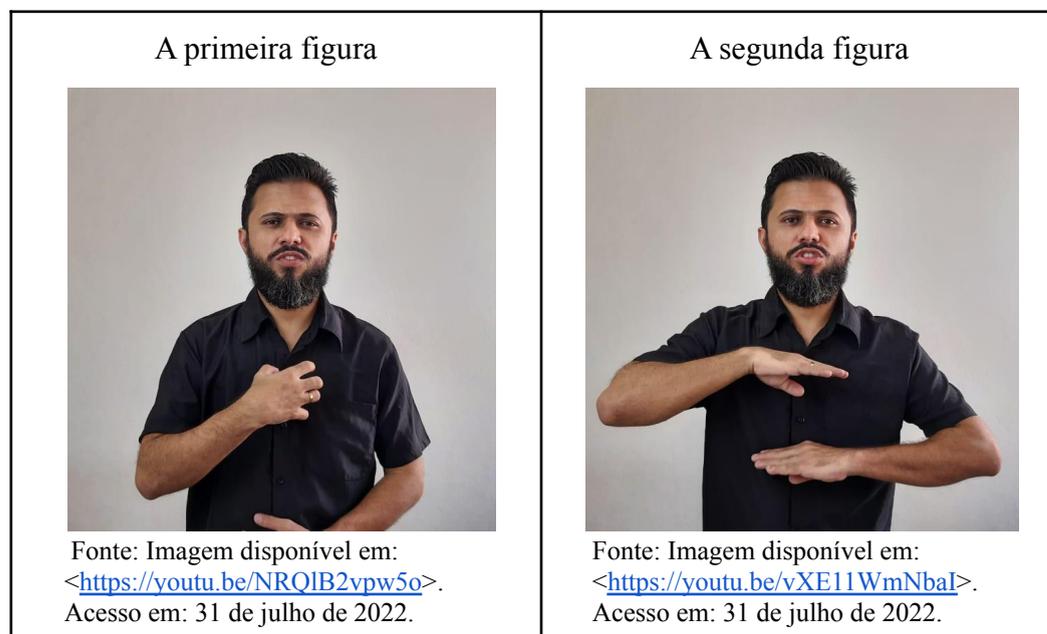
**Figura 28 - Sinal de ENTENDER**

Fonte: Elaborado pelo autor. Imagem disponível em: <<https://youtu.be/n8ncKPhYR8o>>. Acesso em: 31 de julho de 2022.

Desta forma, segundo a transcritora, qualquer alteração fonológica no sinal pode gerar dúvida para o transcritor se ele deve utilizar a mesma glosa ou não. Esta dúvida a cada sinal interfere no tempo de transcrição, tornando o processo todo cansativo e estressante. A transcritora Arapiraca também seguiu este mesmo padrão de pensamento. Ela inicialmente falou das vantagens de utilizar o *SignBank* quando disse “Imagina se o *SignBank* não colocada e assim fica sem auxílio seria pior, como vamos resolver para colocar isto vai ser ruim, por isso é importante ter sinais existentes, é auxílio”. No entanto, ela também falou da necessidade de alimentar o *SignBank* e apresentar alternativas às mudanças fonológicas. Ela citou como exemplo o sinal “JEITO”, explicando que “um sinal para JEITO a mão fica em forma de garra virada no meio do peito com a repetição do movimento, conforme a primeira figura abaixo ou, de outra forma, o sinal de JEITO as duas mãos ficam em forma de “B” com o mesmo movimento, simultaneamente, no meio do peito, conforme a segunda figura

abaixo. É apenas um pouco diferente a configuração da mão, encontrei esse sinal pelo *SignBank* e dispus na transcrição, mas não ficou igual a esse sinal que transcrevi”.

**Figura 29 - Sinal de JEITO**



Fonte: Elaborado pelo autor.

Essa mesma dificuldade foi apontada pelo transcritor Maceió, quando disse que “os sinais de TAMBÉM, com o dedo de indicador em horizontal virado para baixo e o outro, da outra mão, em horizontal virado para cima, e ficam tocando um ao outro com a repetição do movimento, conforme a primeira figura abaixo; e os dedos indicadores juntos, um ao lado do outro, com a repetição do movimento, ver a segunda figura abaixo. Houve um informante que sinalizou o dedo indicador de uma mão parada em boia e uma outra mão fazendo sinais, e que têm diferentes movimentos, se tiver uma mão parada e sem outro apoio, assim o *SignBank* pode interligar tudo na tecnologia,”. Apesar disso, ele ponderou, dizendo que “Por mim melhor é o *SignBank* porque já está tudo pronto sobre isso e disposto, não necessita de ler a regra no manual, também buscar e colocar na transcrição fica mais trabalhoso, pois eu acho que é melhor o *SignBank*, está completo”. Estas considerações sobre o uso do manual e do *SignBank* já apresenta caminhos para entender as duas próximas perguntas, como veremos a seguir.

**Figura 30 - Sinal de TAMBÉM**



Fonte: Elaborado pelo autor.

## 8 - Quais foram os desafios que você encontrou quando estava fazendo as transcrições?

Como dito, muitos problemas já foram mencionados pelos transcritores antes de chegar nesta pergunta. Alguns deles não foram mencionados novamente, outros foram reiterados e ainda outros acrescentados. A transcritora Atalaia apresentou novamente o problema da modalidade entre a Libras e o Português (aqui no formato de glosa), citando novamente o exemplo de PEGAR que sempre vai ser realizado na Libras relacionado ao objeto. Além disso, ela trouxe outras duas dificuldades que ela sentiu no momento da transcrição, como a velocidade de sinalização e os sinais que passavam despercebidos por ela não ter visto utilizando somente uma câmera. Ela explicou que muitas vezes a sinalização era

rápida e não dava para entender e às vezes dava, mas precisava colocar isso, pois, parece uma mosca rápida para perceber, foi difícil para compreender, é preciso ir bem devagar para que possa entender direitinho no momento da produção do vídeo, estava rápido mesmo. [...] toda hora no momento de sinalizar “EU” e não dava para ver os lados e o único jeito de ver é com a câmera de cima, que conseguiu identificar este sinal. Não foi fácil, é necessário usar 4 posições de câmeras mesmo.

Neste relato a transcritora apresentou os problemas e já trouxe as estratégias utilizadas por ela, que, na verdade, são ferramentas e recursos previamente estruturados para os transcritores utilizarem no momento da transcrição. O primeiro é o controle de velocidade do *ELAN*, que permite que configurações de mão, movimentos e pontos de articulação sejam claramente percebidos durante a sinalização. O segundo recurso é a disponibilização de 4 câmeras voltadas para os sinalizantes, o que permite que o transcritor tenha total acesso às configurações de mão e movimentos corporais dos sinalizantes.

Os transcritores Maceió e Arapiraca apresentaram um único problema semelhante: o excesso de tempo despendido para a transcrição. Arapiraca afirmou que “é muito longa para transcrever pois eu aguentava muito, ainda tenho paciência de colocar as glosas, teve demorado bastante e ainda não termina, eu pensei que iria rápido seja possível, mas na verdade não era isso o que eu pensava.”. O trabalho de transcrição, mesmo com as melhores estratégias e recursos tecnológicos sempre vai demandar um tempo considerável do pesquisador. Mas quando questões técnicas e metodológicas embargam este trabalho, o processo se torna ainda mais lento e enfadonho.

Além disso, o grupo de trabalho é fundamental para dividir as tarefas. Quando somente um ou dois transcritores estão trabalhando em um *corpus*, o trabalho tende a sobrecarregar os colaboradores. Maceió fala sobre isso quando relata

eu transcrevi assim mesmo até finalizar, eu ficava tão aliviado, mas, ainda tem mais outro, igual o tempo de 20 min que fiz, eu não faria isso, não. Ainda tenho que fazer outras atividades na Letras-Libras, quer dizer, alguma atividade acadêmica que eu precisava fazer, por isso não consigo me concentrar nessas atividades é impossível para mim.

Como se tratava de um projeto de PIBIC, os alunos tinham outras atividades acadêmicas e o trabalho de transcrição, por ser repetitivo e longo, se torna desinteressante. Nesse contexto, lacunas metodológicas ou problemas técnicos ganham uma proporção muito maior.

## **9 - Quais estratégias você desenvolveu para lidar com estes desafios?**

Como já havia citadas algumas estratégias, Atalaia sintetizou os possíveis caminhos para solucionar os principais problemas apontados por ela. Em primeiro lugar, ela disse que “precisava mais de gravar porque ainda existem sinais de variação, as glosas estão variando bastante, pois é necessário colocar vídeo pelos dados de sinais, porque percebi que esses sinais são de Florianópolis e nem um sinal de Maceió nenhum”. Como a matriz do projeto

nacional é em Florianópolis, muitos sinais que estão no *SignBank* são relacionados a Libras sinalizadas naquela cidade. É importante frisar que o Brasil, por ter proporções continentais, apresenta comunidades surdas com variações lexicais substantivas. Muitos sinais são conhecidos por toda a comunidade surda brasileira, mas cada comunidade utiliza sua própria variante para se comunicar entre si. Desta forma, nós conhecemos os sinais que estão no *SignBank*, mas alguns deles não correspondem às variantes maceioense. A solução para isto, como dito, deve partir do projeto de Maceió, qual seja, alimentar o *SignBank* com os sinais do *corpus* de Maceió.

A segunda proposta diz respeito, não às variações lexicais, mas fonológicas. A transcritora sugere que “se mudou alguma coisa sobre flexibilidade seria ótimo, mas só falta explicar o motivo de flexibilidade sobre sinais e os diferentes movimentos do parâmetro para que saber qual o diferente que tem”. Ela não apresenta exatamente o que deve ser feito, mas aponta que as diferenças fonológicas devem ser levadas em conta de alguma forma pelo *SignBank* para que os transcritores saibam como glosar cada variação.

Os demais transcritores não apresentaram novos caminhos além dos já discutidos nas outras questões, somente explicaram como os desafios da transcrição puderam servir como possíveis temas para as suas pesquisas.

## **10 - Você contribuiu para a criação de novos tokens para o banco da Libras?**

As respostas para esta pergunta foram mais objetivas. Maceió afirmou não ter contribuído com gravações, Arapiraca e Atalaia explicaram que juntas escreveram num caderno todos os sinais que deveriam ser gravados. Além da sugestão de glosa, elas anotaram os sinais em *SignWriting*, como forma de capturar as diferenças fonológicas entre as variantes. No entanto, elas não iniciaram o processo de gravação, por conta da pandemia.

Em suma, percebemos que os problemas apresentados pelos transcritores têm naturezas diferentes, devendo ser, portanto, categorizados e até subcategorizados. Desta forma, conseguimos mapear todos os problemas encontrados de uma maneira mais transparente e, assim, propor soluções abrangentes para cada categoria. No capítulo seguinte, iremos apresentar esta categorização e as possíveis soluções para grande parte destas dificuldades, sendo esta a maior contribuição deste trabalho.

## V. CAPÍTULO - Discussão

Como dito, neste capítulo iremos propor uma categorização dos problemas encontrados no capítulo anterior. Depois disto, iremos apresentar caminhos para que os transcritores, não só deste projeto, mas de todo projeto que se proponha a transcrever *corpus* em língua de sinais possa aplicá-los, testá-los e avaliá-los, contribuindo, assim, para o progresso da linguística de *corpus* de língua de sinais.

### 5.1 Problemas de transcrição

Após observarmos os relatos dos transcritores, percebemos que os problemas não estão somente ligados à tomada de decisão, mas parte da própria estrutura do projeto. Sendo assim, categorizamos os problemas em 3 grupos: problemas sociais, problemas estruturais e problemas metodológicos.

**Tabela 6** - Categorizamos os problemas em 3 grupos

Problemas de transcrição		
Problemas sociais	Problemas estruturais	Problemas metodológicos

Fonte: Elaborado pelo autor.

#### 5.1.1 Problemas sociais

Dentro da questão social, alguns problemas são individuais, pois estão ligados à realidade dos transcritores, outros problemas são locais, pois se relacionam com o curso Letras-Libras de Maceió, e outros problemas são gerais, pois provavelmente se repetiram e se repetem em outros projetos de construção de *corpus* no Brasil. Falaremos brevemente dos dois primeiros, apesar de eles terem uma interferência significativa no processo, e nos ateremos a detalhar os problemas metodológicos, com vistas a construir uma metodologia de transcrição mais precisa para o *corpus* da Libras.

O principal problema social individual encontrado foi: 3 dos 4 transcritores se formaram na fase de transcrição, portanto, não puderam evoluir neste processo. Isto fez com que o quarto transcritor, que não estava nas primeiras fases do projeto, de coleta e tratamento dos dados, assim como não estava nas primeiras formações e discussões, estivesse sozinho

na maior parte do tempo na etapa de transcrição. Apesar do mesmo receber informações pelo manual ou orientações de outros transcritores, sentia-se inseguro para transcrever, situação que ficou clara após o mesmo relatar que ‘eu ficava confuso para usar esse manual ou no *SignBank*, porque nós ainda usamos esse manual até hoje’.

Os problemas sociais estão relacionados diretamente ao contexto alagoano, ao curso Letras-Libras e ao projeto de PIBIC. Em se tratando de Alagoas, o mercado de trabalho para a pessoa surda não é fácil. Desta forma, muitos surdos sequer conseguem estudar no Letras-Libras (por ser um curso matutino), e alguns dos que conseguem, trabalham no turno da tarde, restando uma pequena quantidade de surdos no Curso disponíveis para participar de projetos propostos, passando a ser um problema geral para o Curso, que conta com alunos surdos, não só no PIBIC, mas em outros projetos como a Casa de Cultura no Campus (voltado para o ensino de Libras) e o Projeto TATU (voltado para acessibilidade e tecnologia).

Somado a isto, o projeto do PIBIC contou apenas com uma bolsa de estudos, inviabilizando a participação de alunos que tinham vontade de participar do projeto, mas não possuíam o mínimo de recurso financeiro para passar o dia inteiro na Universidade, por exemplo. Esse contexto social local pode ter impedido a participação de mais surdos no projeto, o que, por sua vez, fez com que somente 4 transcritores, em momentos diferentes, trabalhassem na transcrição do *corpus*.

Por fim, os problemas sociais gerais envolvem toda a comunidade surda. É sabido que o sistema educacional brasileiro não oferece uma educação de qualidade adequada às especificidades linguísticas da comunidade surda. Desta forma, muitos surdos, mesmo estando no nível de graduação, apresentam dificuldade de compreensão da língua portuguesa, em diferentes níveis, o que torna o processo de glosagem mais lento. Como relatado, a diferença de modalidade por si só já torna o trabalho de glosagem mais lento e complexo, mas quando não há um domínio por parte do transcritor da língua da glosa, o trabalho tende a ser mais demorado, ainda que haja o auxílio do *SignBank*.

### 5.1.2 Problemas estruturais

Nós pensamos numa categorização que englobasse questões relacionadas à prática de glosagem, mas que não estão diretamente relacionadas às decisões que precisam ser

tomadas. Ao mesmo tempo, são problemas que estão em certa medida ligados aos problemas sociais, sendo assim, questões mistas, que envolvem metodologia e contexto social.

A primeira questão a ser observada é a prática de transcrever no *ELAN*. A carência de prática no *ELAN* pode fazer com que o transcritor leve um tempo considerável para transcrever. Outra questão é o conhecimento metalinguístico para transcrever. É importante que os transcritores não só conheçam a Libras e o português, mas saibam segmentar os sinais, percebendo as fronteiras entre os sinais, as questões morfológicas e sintáticas que influenciam na fonologia do sinal em dados contextos. Além disso, para marcar as DVs, é preciso saber o que de fato é uma DV e o que é um sinal. Para marcar um IX é preciso saber o que é uma marcação pronominal na Libras, e para marcar um E, o que é gesto (emblema) também é necessário saber o que é gesto. Todas estas peculiaridades exigem uma formação e uma prática intensiva por parte dos transcritores antes mesmo de iniciar a transcrição de fato. Ainda que os transcritores tenham participado de formações e tenham recebido instruções sobre como interpretar o manual, vimos que os transcritores apresentaram dificuldades em relação a estas questões metalinguísticas e técnicas, o que pode ser explicado pelas questões sociais já mencionadas na seção anterior, como as demais atividades acadêmicas, o baixo número de bolsas e o fato de alguns colaboradores terem se formado na etapa da transcrição, o que fez com que este processo de formação fosse interrompido por vezes.

Havia um grupo de trabalho inicialmente no projeto, mas na etapa da transcrição, muitos saíram do projeto e outros se formaram. Estes desfalques no grupo de colaboradores para discutir em conjunto as dificuldades de transcrição de cada um, tornaram o trabalho mais difícil, pois muitas questões básicas de um ou outro transcritor podem ser sanadas por um outro que saiba como resolvê-las. Em suma, os problemas sociais acarretaram estes problemas estruturais e, por sua vez, estes problemas estruturais poderiam evitar muitos problemas metodológicos.

### 5.1.3 Problemas metodológicos

Aqui desenvolvemos a maior parte desta discussão. Levando em conta todo o contexto que envolve o processo de transcrição, iremos nos debruçar sobre as questões que podem ser resolvidas de maneira técnica e que envolvem a tomada de decisão do transcritor.

Ou seja, estes problemas estão para além do contexto social e estrutural do projeto e, quando somados a estes, tornam o trabalho quase que inviável.

Nós subcategorizamos os problemas metodológicos em 4 contextos, a saber: o efeito de modalidade, a variação fonológica, a variação lexical, e a limitação vocabular do *SignBank*. Apresentaremos os contextos, trazendo exemplos citados pelos transcritores.

**Tabela 7** - Subcategorizamos os problemas metodológicos em 4 contextos

Problemas metodológicos			
Efeito de modalidade	Varição fonológica	Varição lexical	Limitação vocabular do <i>SignBank</i>

Fonte: Elaborado pelo autor.

#### 5.1.3.1 O efeito da modalidade

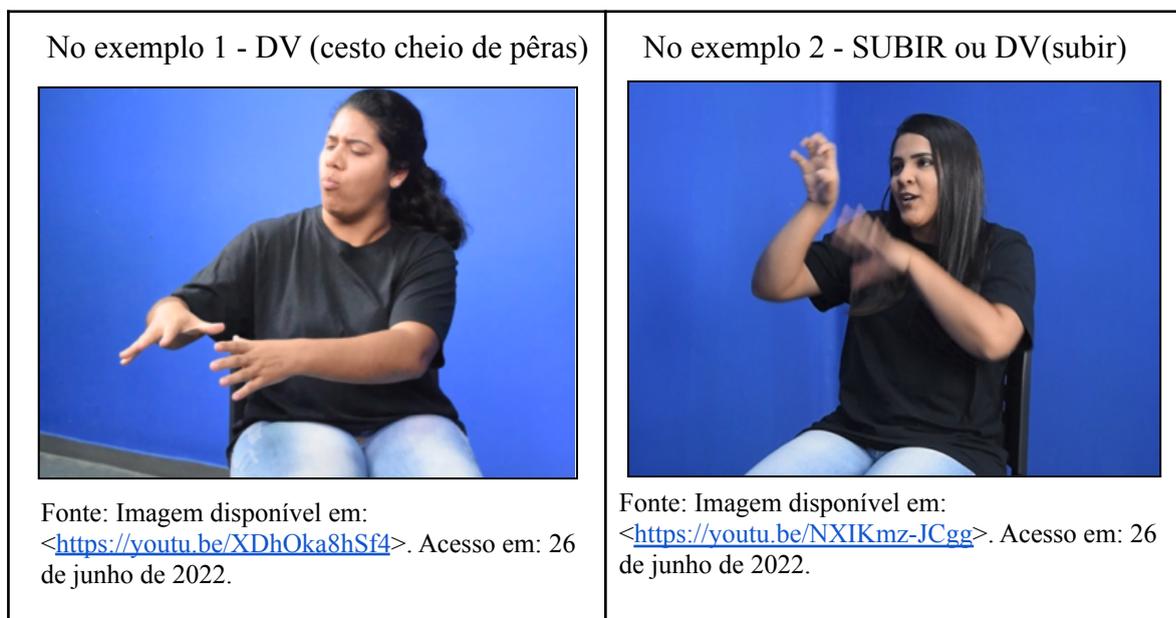
O primeiro contexto é quando as duas línguas (a Libras e o Português), por conta da modalidade, apresentam estruturas absolutamente distintas. A despeito de alguns contextos sintáticos específicos, as maiores diferenças de produção linguística acontecem no uso de descrições visuais (DVs) ou classificadores na Libras, os predicados com classificadores organizam a estrutura sintática de uma maneira não-linear, pois muitas vezes verbo e objeto, ou sujeito e verbo, estão sendo realizados num mesmo sinal.

Além disso, se já entendemos que, mesmo num processo como este de glosagem, temos dificuldade para encontrar equivalentes semânticos em duas línguas diferentes (de modalidades diferentes), pois cada língua estrutura seu léxico e seus significados de maneira distinta, os classificadores são sinais que não estão no núcleo do léxico, portanto, não apresentam um significado estável, mas abertos a mudanças de significados de acordo com a combinação com outros classificadores ou sinais.

Este problema aparece, tanto quando o transcritor tem que pensar no significado do predicado, quanto quando ele tem que pensar numa palavra no português que expresse um significado mais ou menos equivalente. Ainda que já tenha sido decidido que as descrições visuais não precisam ser detalhadas, há sinais estáveis que são oriundos de classificadores. E há sinais em que o transcritor não consegue decidir se é um sinal estável ou uma DV. Nos exemplos a seguir, vemos estas diferenças. No exemplo 1, vemos um classificador de forma indicando o quão cheio a cesta de frutas estava, portanto é uma descrição visual que apresenta a forma do objeto. Já no exemplo 2, alguém está subindo uma escada. O

classificador aqui é uma ação que poderia ser glosada como subir, ou somente como DV (nas legendas colocaremos a ação da DV entre parênteses somente para ilustração, pois como já dito, ela não é mais necessária na transcrição).

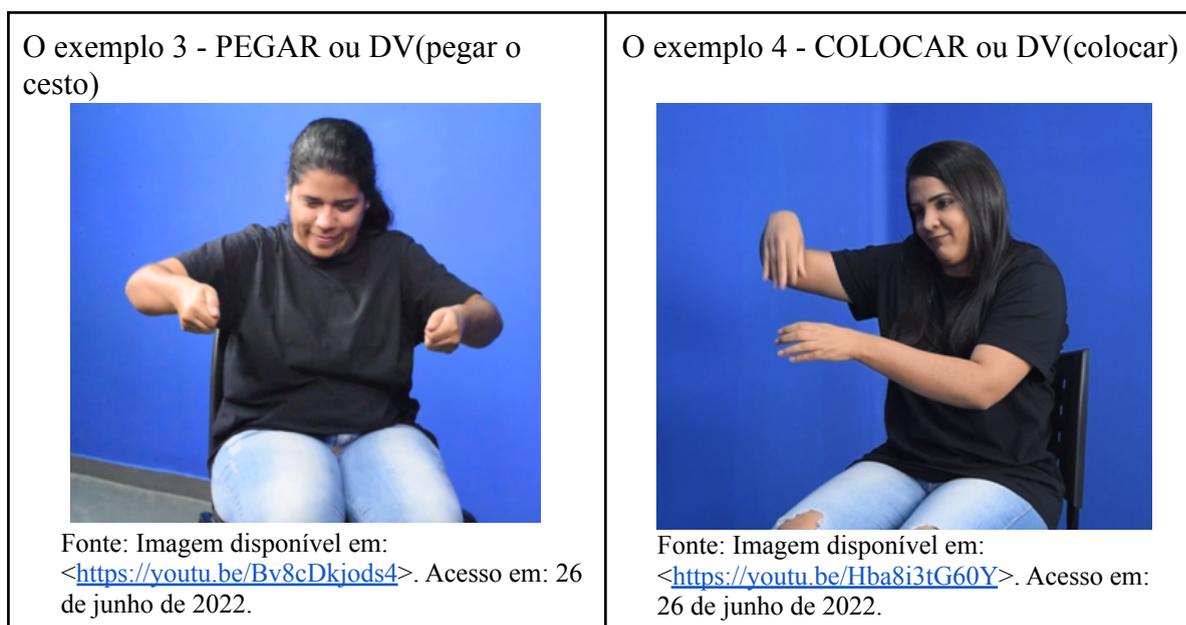
**Figura 31** - Descrição visual (DV)



Fonte: Elaborado pelo autor.

Os exemplos 3 e 4 trazem os verbos PEGAR e COLOCAR, que, como já explicado, sempre poderão se transformar a depender do objeto.

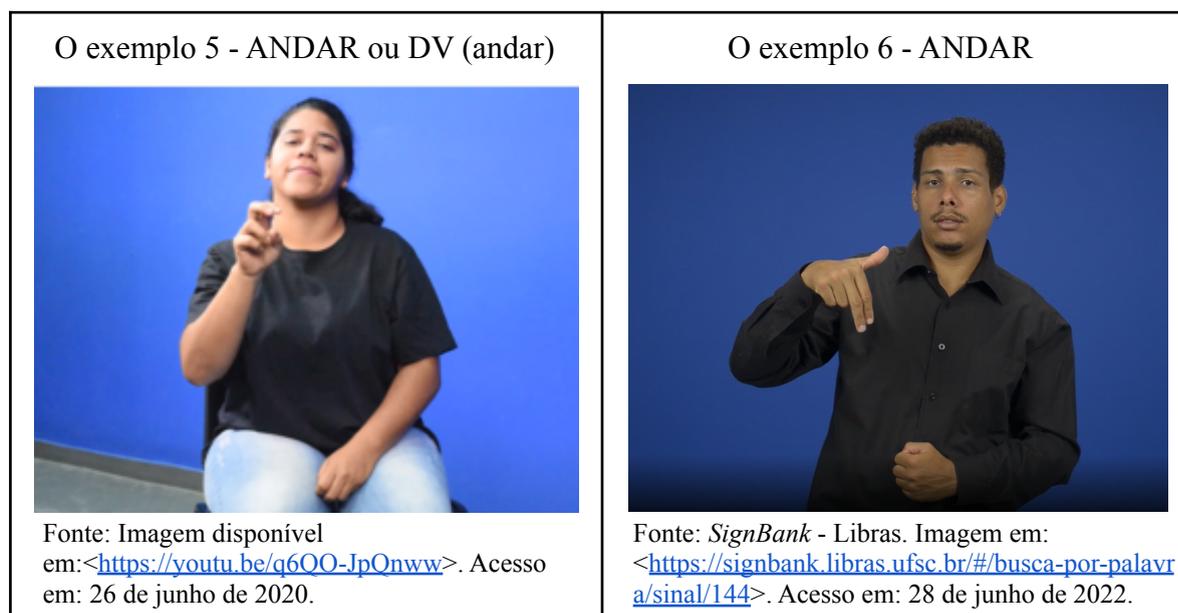
**Figura 32** - Sinal de PEGAR e COLOCAR



Fonte: Elaborado pelo autor.

É possível que tanto a glosa DV, quanto o sentido semântico do classificador seja glosado e isso gerou dúvidas entre os transcritores. Além disso, no caso do exemplo 5, o sinal ANDAR é feito como uma DV, pois envolve toda expressão corporal do surdo, mas ao mesmo tempo é um sinal comumente usado na Libras sinalizada em Maceió (mesmo não constando no *SignBank*). O sinal que consta no *SignBank* apresenta uma configuração de mão e movimento diferentes, conforme vemos no exemplo 6.

**Figura 33** - Sinal de ANDAR



Fonte: Elaborado pelo autor.

Como vimos, não há uma relação de um para um entre sinais e palavras no português nestes exemplos e, ainda, as orientações do manual não contemplam todas as possibilidades de descrições visuais na Libras. Um outro recurso ainda não mencionado que surgiu como um problema de efeito de modalidade durante a transcrição foi o uso de “boia” no *corpus* (ver Heitkoetter & Xavier (2021)). A boia é quando um sinal é suspenso com uma mão no espaço de sinalização e a outra mão continua sinalizando em relação à primeira mão. É muito comum em Libras utilizarmos bóias de listagem (ver Heitkoetter & Xavier (2021)), quando uma mão é suspensa com os dedos selecionados representando os números e a segunda mão continua sinalizando fazendo referência a cada dedo selecionado.

Apresentaremos agora um exemplo claro deste recurso. Na sequência abaixo, a sinalizante mantém com a mão esquerda a configuração do número 3, representando 3 entidades. Ela então aponta para cada dedo e realiza um sinal qual seja uma característica de

cada entidade. Por fim, ela segura dois dos três dedos para dizer algo sobre a segunda e a terceira entidade.

**Figura 34** - Representando 3 entidades



Fonte: Elaborado pelo autor.

Este formato de construção é muito produtivo em Libras e só é possível porque há dois articuladores nesta modalidade de língua. Com isto, há uma simultaneidade de informações que faz com que a transcrição para o português se torne quase inviável. Ainda que já haja uma trilha para cada mão, a mão boia não é o problema, mas a mão principal que está sinalizando produzindo sinais em função da boia, principalmente quando observamos a última imagem da sequência, que representa DOIS DOS TRÊS.

Por fim, apresentaremos exemplos do uso de gestos no *corpus*. Na figura abaixo, vemos três exemplos. Por mais que haja uma indicação no manual para o uso de um código específico para a transcrição de gestos (emblema) (a saber, “E” (descrição semântica do sinal)), os transcritores tiveram dificuldades em identificar alguns sinais como gesto (emblema), conforme a figura abaixo.

**Figura 35** - Gesto (emblema) com código E



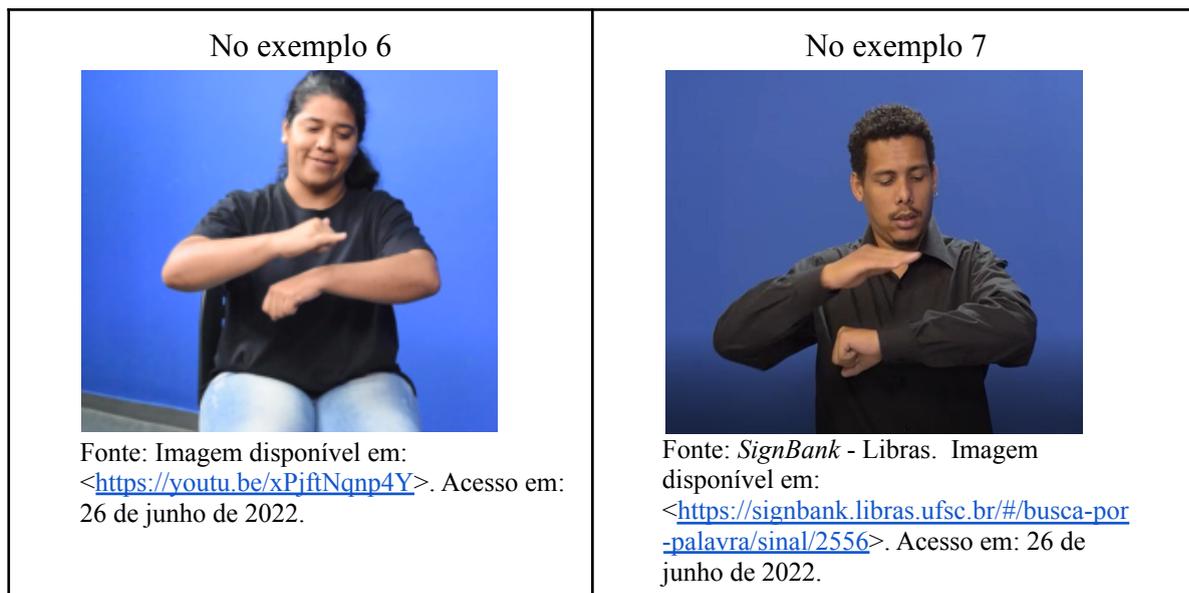
Fonte: Elaborado pelo autor.

A legenda dos exemplos representa a glosa proposta pelo transcritor. É notável que há uma dúvida em relação à semântica do sinal. Na entrevista, o transcritor nos afirmou não saber qual glosa seria ideal para estes sinais.

### 5.1.3.2 Variação fonológica

A variação fonológica foi uma das principais dificuldades encontradas pelos transcritores. Isto se deve a alguns fatores. Em primeiro lugar, porque não há no *SignBank* a possibilidade de marcar essas diferenças que são, em certos momentos, muito notáveis. Em segundo lugar, porque ela pode acontecer em níveis mais sutis, ou mais perceptíveis, portanto, se formos levar todas as mínimas diferenças de realização de um sinal em conta, teremos que lidar com um número expressivamente maior de glosas. Seria difícil decidir quais diferenças de realização devem ser marcadas como uma nova glosa e quais são somente diferenças pragmáticas.

No exemplo 6, a sinalizante realiza o sinal SAPO com o polegar e o dedo mínimo selecionados (abertos). No exemplo 7, no *SignBank* o sinal é realizado com todos os dedos selecionados e juntos, conforme imagem abaixo, configurando assim uma variação fonológica de configuração de mão.

**Figura 36 - Sinal de SAPO**

Fonte: Elaborado pelo autor.

O exemplo a seguir nos mostra mais um exemplo de variação fonológica, mas desta vez, a própria sinalizante utiliza as duas formas. Desse modo, não podemos dizer que se trata de duas variantes regionais, mas ainda assim, sendo uma variação fonológica, trouxe problemas para a transcrição. Uma das variantes está no *SignBank*, com a mesma configuração de mão no formato do sinal da letra A, mas a outra variante com a configuração de mão do sinal da letra S não consta no *SignBank*.

**Figura 37 - Sinal de ANSIEDADE**

Fonte: Elaborado pelo autor.

No sinal abaixo, vemos mais uma variação fonológica. A configuração da mão do sinal feito pela participante do *corpus* seleciona todos os dedos, enquanto que a configuração de mão do sinal do *SignBank* seleciona somente o dedo indicador.

**Figura 38** - Sinal de ALIVIAR



Fonte: Elaborado pelo autor.

Como vemos, as diferenças entre as configurações de mão não são expressivas, vão além do que um simples desvio na sinalização, pois em outros sinais, estas mudanças seriam distintas. É claro que não há mudança no significado dos sinais, portanto, são alofones, mas a questão que se coloca é se estas diferenças devem ser simplesmente ignoradas ou transcritas de forma diferente.

Apresentaremos agora dois exemplos que nos ajudaram a refletir de maneira mais profunda sobre as variações fonológicas no processo de transcrição. Nos exemplos a seguir, não há simplesmente uma variação, mas o significado varia. Vejamos os exemplos.

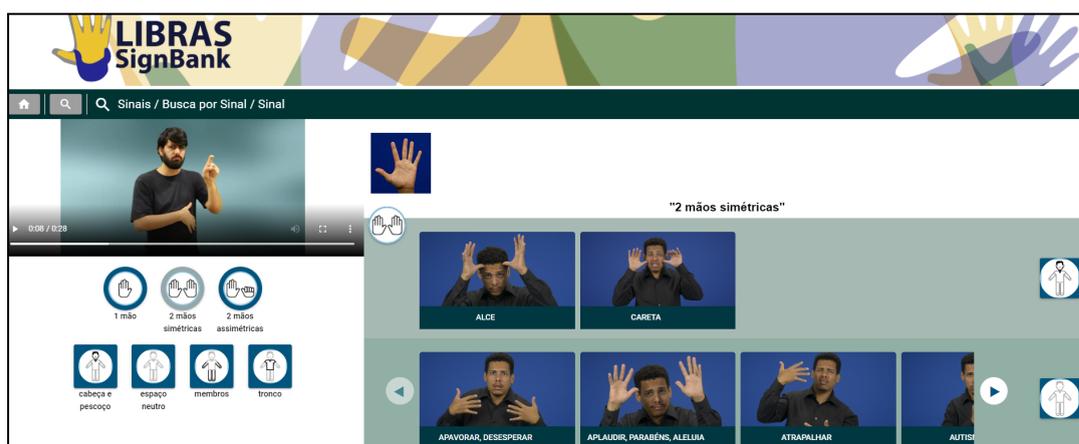
**Figura 39 - Sinal de ALCE e VEADO**



Fonte: Elaborado pelo autor.

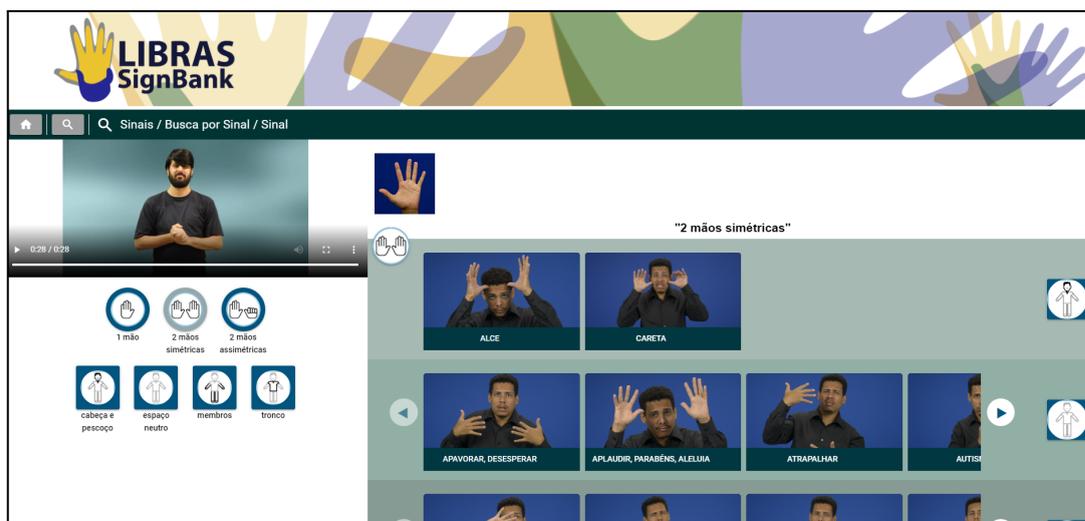
Quando o transcritor procurou pelo sinal VEADO no *corpus*, encontrou outra variante, e apresentou o problema glosando o sinal com XXX. Este seria um caso como os demais de variação lexical, pois há uma mudança na configuração de mão no começo do sinal e um movimento para cima. No entanto, ao procurarmos no *SignBank*, não pela palavra, mas pelo filtro da configuração de mão (ver figuras 41 e 42), encontramos um sinal idêntico, mas tendo como glosa ALCE.

**Figura 40 - Configuração de mãos**



Fonte: SignBank - Libras. Imagem disponível em: <https://signbank.libras.ufsc.br/#/sinais/configuracao-de-mao>. Acesso em: 27 de junho de 2022.

**Figura 41** - 2 mãos simétricas



Fonte: *SignBank* - Libras. Imagem disponível em: <https://signbank.libras.ufsc.br/#/sinais/configuracao-de-mao/buscar-sinal/7>. Acesso em: 27 de junho de 2022.

O mesmo aconteceu com os sinais ABELHA e MOSQUITO. O sinal ABELHA realizado pela participante do *corpus* se difere do sinal ABELHA do *SignBank* em relação ao ponto de articulação final, pois na variante da participante o ponto final é na mão de apoio, enquanto que na variante do *SignBank* o ponto é na bochecha. Entretanto, ao procurarmos pelo sinal utilizando o filtro da configuração de mão, encontramos um sinal idêntico ao da participante, mas glosado como MOSQUITO.

**Figura 42** - Sinal de MOSQUITO e ABELHA



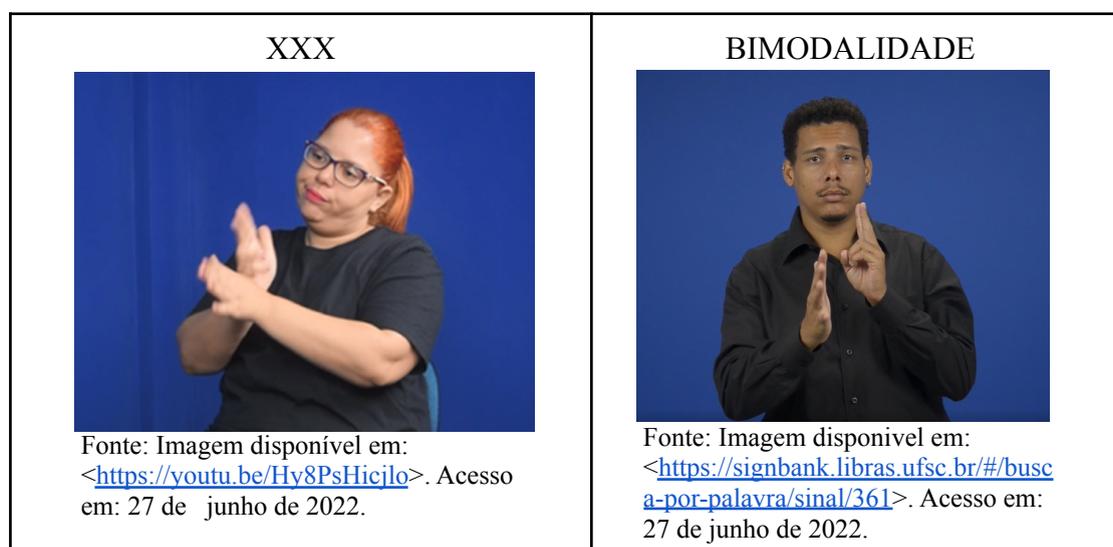
Fonte: Elaborado pelo autor.

Iniciaremos esta parte apresentando o sinal BIMODALIDADE. Este sinal se refere a um modelo de educação de surdos do século passado, portanto, foi um sinal criado em acordo entre os surdos e difundido para todo o Brasil. O transcritor procurou pelo referido sinal no *SignBank* e não o encontrou, ao menos do modo como fora realizado pela sinalizante, motivo pelo qual foi marcado com XXX em vez da glosa. No entanto, no banco de sinais encontramos, quando a análise, uma variante do sinal, como se pode ver em 43 (figura do lado esquerdo). O sistema de glosa, embora bastante utilizado por pesquisadores de línguas de sinais, em certa medida limita as análises, sobretudo se o transcritor não conhecer variantes do sinal.

Os sinais abaixo representam BIMODALIDADE. No *Corpus* o encontramos com configuração de mão dominante com os três dedos para cima - polegar, indicador e médio da mão direita - e dois dedos fechados - anelar e mínimo. Já a mão não dominante, encontra-se em posição contralateral, aberta, abaixo da mão dominante. A variação fonológica encontrada no banco de sinal pode ser descrita da seguinte forma: mão direita aberta e a esquerda em V, com movimentos sincrônicos alternados, com ambas as mãos em dominância, no espaço neutro, à frente do sinalizante.

O *SignBank* auxilia bastante no momento da transcrição dos dados, porém é necessária atualização constante, incluindo-se aí glosas que podem ser cognatas. Buscou-se, por exemplo: BIMODALISMO, BIMODAL, mas sem sucesso, tendo encontrado apenas BIMODALIDADE. Importante ressaltar que o transcritor necessita, também, conhecer a glosa para o sinal. Do contrário de nada adianta o banco de sinais, o *SignBank*.

**Figura 43** - Sinal de **BIMODALIDADE**



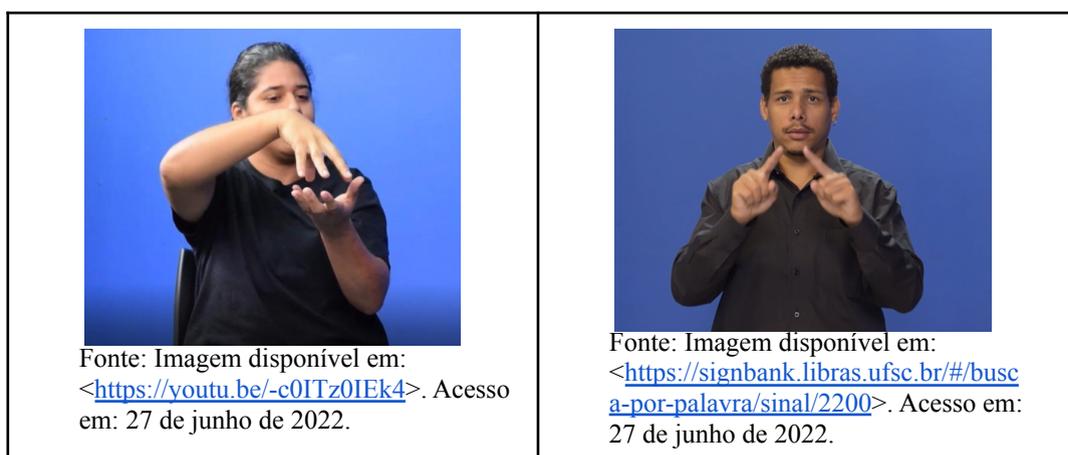
Fonte: Elaborado pelo autor.

Em suma, percebemos que há na verdade dois problemas relacionados à variação fonológica. Um problema é a importância de registrar em glosa estas diferenças fonológicas, por não serem meras diferenças pragmáticas. O segundo problema é que a pesquisa por configuração pode identificar um sinal homônimo ao sinal realizado pelo participante do *corpus*, ou seja, um sinal com uma fonologia idêntica, mas com significados diferentes. Nós sabemos que um sinal com significados que se modificam a depender do contexto devem receber uma mesma glosa padrão, que consta no *SignBank*. Mas aqui, o significado muda por ser um sinal de uma região diferente, independente do contexto linguística. Podemos, portanto, interpretar estas diferenças como uma variação lexical, se compararmos o sinal para VEADO que consta no *SignBank* com o sinal do *corpus*, ou uma homonímia se ignorarmos a diferença regional e nos atermos somente aos sinais VEADO(*corpus*) e ALCE(*SignBank*). A tomada de decisão para caracterizar a diferença, se homonímia ou variação, torna o trabalho mais complexo e lento.

### 5.1.3.3 Variação lexical

Apresentaremos agora alguns exemplos de variação lexical no *corpus*. A partir destes exemplos, iremos entender como isto pode ser um problema para a transcrição. Iniciaremos com o sinal PÊRA. O sinal realizado pela participante do *corpus* é uma variante regional não registrada no *SignBank*, e a variante do *SignBank* não é sinalizada em Maceió. Entendemos que não há como uma pessoa ou um só grupo de uma determinada região alimentarem o *SignBank*, mas este trabalho deve envolver todas as variantes da Libras do Brasil.

**Figura 44 - Sinal de PÊRA**



Fonte: Elaborado pelo autor.

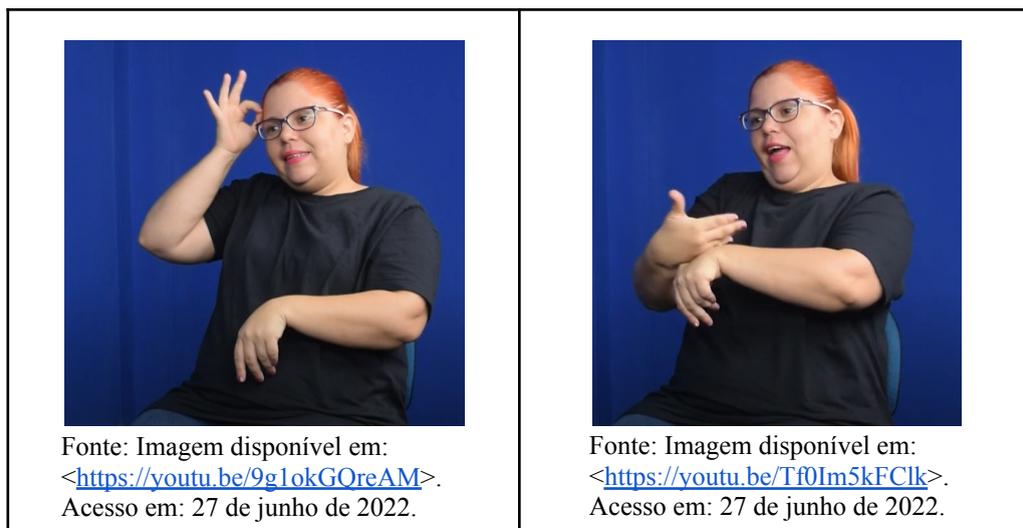
Já para este próximo exemplo, vemos que a variação lexical é prevista no *SignBank* para o sinal CONCURSO, mas ainda assim, não contemplou a variante utilizada pela participante do *corpus*.

**Figura 45 - Sinal de CONCURSO**



Fonte: Elaborado pelo autor.

Por fim, iremos apresentar o sinal CEDO. A figura abaixo apresenta duas variantes para este sinal. Percebe-se que mais uma vez a mesma participante realiza as duas variantes, portanto, não se trata de uma variação regional, mas ainda assim é uma variante lexical. O curioso neste caso é que nenhuma das variantes se encontra no *SignBank*, tampouco há outra glosa para a semântica de CEDO. Este caso apresenta tanto um problema de variação lexical, quanto um problema de limitação do *SignBank*, foco da próxima parte desta seção.

**Figura 46 - Sinal de CEDO**

Fonte: Elaborado pelo autor.

Entendemos aqui que há variações lexicais que estão dentro de uma comunidade ou que variam de uma região para outra. Há ainda variações que são regionais e locais para um mesmo referente. A quantidade de variações lexicais na Libras é expressiva, principalmente quando levamos em conta todo o território nacional. Desta forma, é esperado que as variações lexicais previstas no *SignBank* não sejam suficientes para dar conta de todas as possibilidades de sinalização do léxico da Libras.

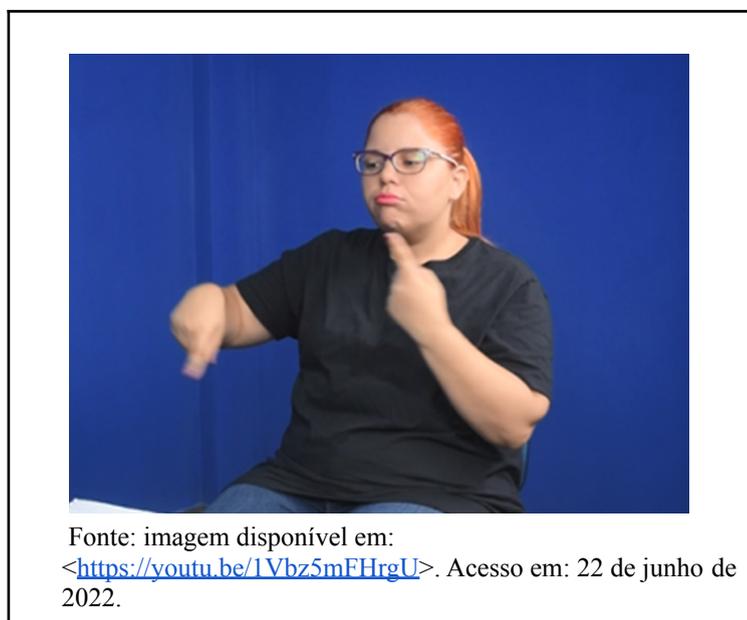
#### 5.1.3.4 Limitação vocabular do *SignBank*

Agora iremos apresentar alguns itens lexicais que não apresentam uma variação lexical, pois seus referentes sequer estão registrados no *SignBank*. Isto acontece porque o léxico de uma língua como a Libras (com abrangência em todo o território brasileiro, presente em diversos ambientes sociais) possui um extenso léxico, que se aumenta a cada dia. Desta forma, um *SignBank* sempre vai apresentar uma limitação quanto ao registro do léxico de uma língua.

Este segundo exemplo é um sinal que surgiu como um classificador, mas que é muito comumente utilizado. Entendemos que se trata de um sinal lexicalizado, com um significado e forma estáveis. Desta forma, não entendemos que a glosa correta seria DV, mas uma glosa específica. Entretanto, o sinal não está presente no *SignBank*. Não é possível saber se este sinal tem uma abrangência nacional, mas é comum pensar que sim, pois apresenta uma

forma gestual icônica forte. É possível, no entanto, que ele seja facilmente entendido, mas não comumente produzido em outros lugares.

**Figura 47 - VOLTAS E VOLTAS**

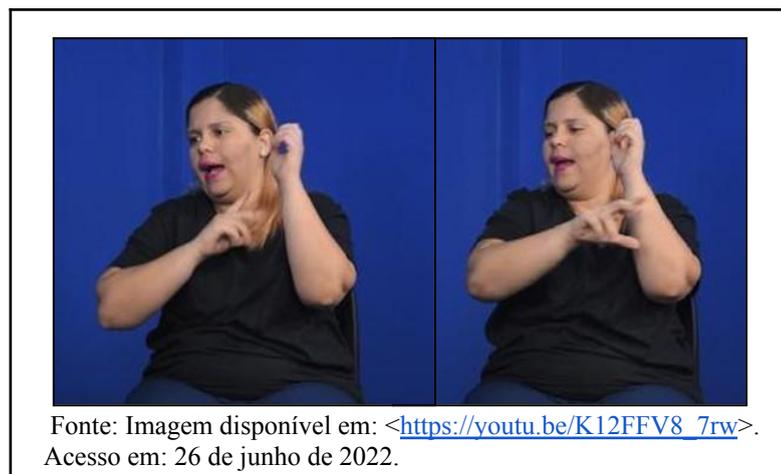


Fonte: Elaborado pelo autor.

Por fim, o sinal DIVERSIDADE DE ÁRVORES. Adotamos este nome somente para a análise, pois ele ainda não está glosado no *corpus*, nem no *SignBank*. O sinal é uma mescla de dois outros sinais, ÁRVORES e DIVERSAS feitos simultaneamente, de forma que o sinal DIVERSAS apresenta um movimento que parte da mão que está sinalizando ÁRVORE para fora, ou seja, o sinal DIVERSAS tem o sinal ÁRVORES como ponto de articulação inicial. É um sinal que apresenta uma boa formação, pois a mão de apoio ÁRVORE está sem movimento, enquanto que a mão principal se movimenta.

Pensamos se este sinal não poderia ser glosado simplesmente com duas glosas separadas, mas é evidente que se trata de um único sinal. Este sinal não se repete no *corpus*, mas desafia a transcrição, pois nos mostra que, como este, muitos outros sinais podem ser oriundos de um processo de composição na Libras por surdos na comunidade surda todos os dias, tornando o processo de glosagem ainda mais complexo.

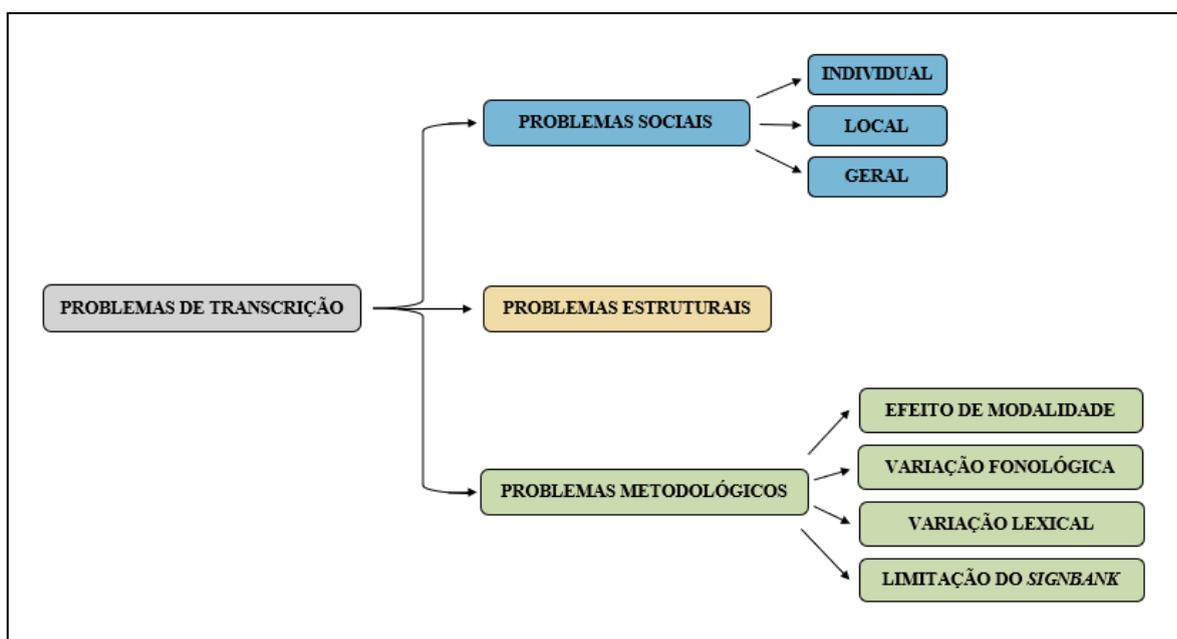
**Figura 48 - ÁRVORE DIFERENTES**



Fonte: Elaborado pelo autor.

O esquema abaixo apresenta um resumo com os problemas de transcrição explanados nesta seção.

**Esquema 1 - Os problemas de transcrição**



Fonte: Elaborado pelo autor.

Na seção seguinte iremos apresentar quais possíveis caminhos podem ser percorridos para que estes problemas se dissolvam, ou que sejam administrados de maneira que a transcrição não seja prejudicada.

## 5.2 Caminhos para solucionar os problemas de transcrição

Nesta etapa, iremos propor caminhos para solucionar os problemas descritos na seção anterior. É importante salientar que não tivemos tempo para testar a transcrição a partir das propostas que virão, portanto, não podemos chamar de solução, mas de caminhos para a solução, que vai exigir testes e mais estudos a fim de encontrar uma maneira produtiva de realizar as glosagens de *corpus* de língua de sinais. Inicialmente iremos abordar os problemas sociais e estruturais, e, em seguida, os problemas metodológicos.

### 5.2.1 Caminhos para os problemas sociais

Para discutir os problemas sociais, vale frisar que as problemáticas apresentadas são de ordem política e que na maior parte das vezes extrapolam o âmbito da pesquisa, a começar do fato de que a entrada de surdos na Universidade é baixíssima, e isto se deve a um sistema educacional defasado que não prepara as crianças e adolescentes surdos para a educação superior. Para que pesquisas de qualidade na área da Libras sejam realizadas com um protagonismo surdo, é mister oferecer aos surdos boas condições para que eles possam adentrar a universidade.

Mas não só isso, é importante estabelecer políticas estudantis que garantam que estes surdos irão ter assistência para ir à Universidade, acompanhar as leituras, e participar de projetos de extensão e pesquisa. Para tanto, os projetos devem oferecer bolsas de estudo condizentes com a realidade atual do aluno universitário, que precisa de recursos financeiros para se alimentar, se locomover e estudar.

Após garantir a entrada e permanência de alunos surdos na Universidade, é preciso que os alunos que desejem fazer pesquisa possam ter contato com a pesquisa ainda no início da graduação, como aconteceu comigo. Desta forma, os alunos irão ter tempo para participarem de formações e assim os projetos de pesquisas poderão sempre contar com um grupo de colaboradores que se apoiarão durante toda a pesquisa.

### 5.2.2 Caminhos para os problemas estruturais

Minimizando os problemas sociais, o projeto poderá criar uma estrutura que contemple as necessidades dos colaboradores surdos e as particularidades da pesquisa com a

Libras. Com um grupo de colaboradores, além das formações em outras universidades e com outros professores, após certo tempo de experiência, os colaboradores poderão eles mesmos acompanhar os novos alunos no uso do *ELAN*, do manual e do *SignBank*. Eu percebi que algumas questões levantadas pelos colegas transcritores poderiam ser resolvidas por mim, pois eu participei de outras formações anteriormente e tive contato com discussões que envolviam problemas parecidos.

O grupo poderia também revisar os trabalhos uns dos outros, pois a dificuldade de um transcritor pode não ser a mesma que a do outro. Além disso, um só transcritor, quando não está 100% certo de uma determinada glosa, tende a não preenchê-la. No entanto, quando há um grupo realizando o trabalho, os componentes podem avaliar e validar a decisão do transcritor. Discutir a melhor maneira de utilizar o manual e o *SignBank* também deve ser feito em grupo. Vale destacar que esta estrutura deve estar presente em todo o processo de construção do *corpus*, não só na etapa de transcrição, mas na coleta e tratamento dos dados. Desta forma, todos os transcritores adotarão a mesma metodologia de transcrição, evitando assim glosagens diferentes.

Ainda sobre as questões estruturais, os colaboradores precisam de formações específicas na área da linguística, pois muitas vezes as disciplinas de linguística na graduação não contemplam todo o conhecimento metalinguístico necessário para que o transcritor possa realizar a transcrição com segurança. Não podemos somente dizer para o transcritor marcar as descrições visuais com DV, mas explicar no manual de transcrição detalhadamente o conceito de DV adotado para o *corpus*, quais características devem ser encontradas para que o transcritor classifique o sinal como DV e não como um sinal estável. É importante lembrar que a transcrição já é uma espécie de análise linguística, ainda que se proponha a ser mais abrangente.

Caso dúvidas e novos problemas metodológicos surjam, de forma que o grupo de trabalho não consiga chegar a um consenso, ou não saiba como resolver, uma possibilidade seria estabelecer encontros entre projetos de construção de *corpus* no Brasil. Nestes encontros, os colaboradores podem apresentar os problemas encontrados para que os alunos e professores dos outros projetos possam contribuir e, além disso, trocar experiências para otimizar ainda mais o processo de transcrição. O projeto matriz da Universidade de Santa Catarina tem auxiliado os demais, inclusive o de Maceió durante todo o processo de construção do *corpus*, assim como os demais projetos. No entanto, com muitas dúvidas e questões, os encontros não dão conta de sanar todas as dificuldades. Desta forma, esta prática de continuar existindo em consonância com as demais estratégias pensadas aqui.

### 5.2.3 Caminhos para os problemas metodológicos

Após estabelecer uma estrutura organizacional adequada para o trabalho de transcrição, é necessário agora pensar em soluções metodológicas que facilitem o processo de glosagem. Dividiremos os caminhos em quatro partes, uma parte para cada tipo de problema listado.

#### 5.2.3.1 Caminhos para os problemas de efeito de modalidade

Ao entendermos que a língua portuguesa, na qual a glosa vai ser escrita, possui uma modalidade diferente da Libras, se faz necessário levar em conta cada diferença de modalidade no momento da glosa. Como descrevemos, há sinais que são classificadores e, por isso, não possuem forma e significado estáveis; há sinais que são uma mescla de classificador com um sinal estável, há gestos e há ainda as boias que fazem parte da modalidade visual e não estão presentes na língua portuguesa.

Desta forma, precisamos pensar em soluções para cada uma destas diferenças. Vale salientar que o manual já apresenta uma significativa contribuição para este entendimento, e, como já dito, é mister que os transcritores tenham domínio dos conceitos trazidos pelo manual para que possam tomar decisões padronizadas no momento da transcrição. Entretanto, como percebemos, apesar da contribuição do manual, algumas questões se mantêm no momento da transcrição.

No manual mais recente, a orientação sobre o uso de DV se resume a seguinte explicação:

**Figura 49** - Descrição visual - DV

Os chamados verbos “classificadores” são referidos como verbos descritivos visuais. Para os eventos descritos será usada a sigla DV sem que haja necessidade de descrever o evento. Por exemplo:

00:00:59.000	00:01:00.000	00:
DV( <del>pés-andando-em-chão</del> )		E(gorc)
DV( <del>pés-andando-em-chão</del> )		E(gorc)

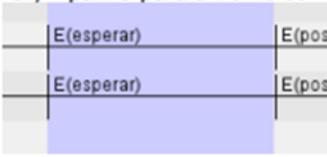
Fonte: Manual de transcrição do *Corpus LIBRAS* Inventário Nacional (versão 19).

A explicação é útil, mas não explica a diferença entre sinal, gesto(emblema) e DV. É mister estabelecer distinções entre estas possibilidades no manual para que o surdo tenha total clareza de que caminho deve escolher para transcrever. Para a transcrição de Gestos(emblema), o manual orienta que a transcrição seja feita conforme a figura abaixo. O código a seguir está na versão de 2017, mas não consta na versão de 2019, mesmo que a orientação tenha sido mantida até hoje.

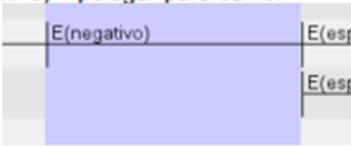
**Figura 50 - Gesto (emblema ‘E’)**

A produção de emblemas que são altamente convencionalizados será transcrita observando os Identificadores de Sinais:

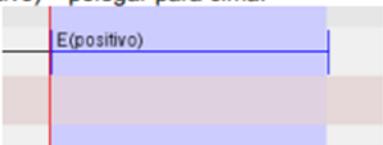
**E(esperar)** – palma para a frente com CM em B, movimento leve e repetido.



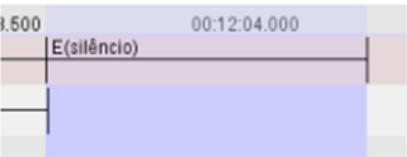
**E(negativo)** – polegar para baixo.



**E(positivo)** – polegar para cima.



**E(silêncio)** – dedo indicador nos lábios.



Outros exemplos:



A produção de emblemas que são altamente convencionalizados (lexicalizados) será transcrita observando-se os sinais da Libras *SignBank*. Assim, sinais como ENTÃO, ESPERAR, SILÊNCIO passam a ser transcritos como glosas do Português normalmente. Se o emblema não constar no Libras *SignBank*, deve ser incluído pelo transcritor.

Fonte: Manual de transcrição do *Corpus LIBRAS Inventário Nacional* (versão 19).

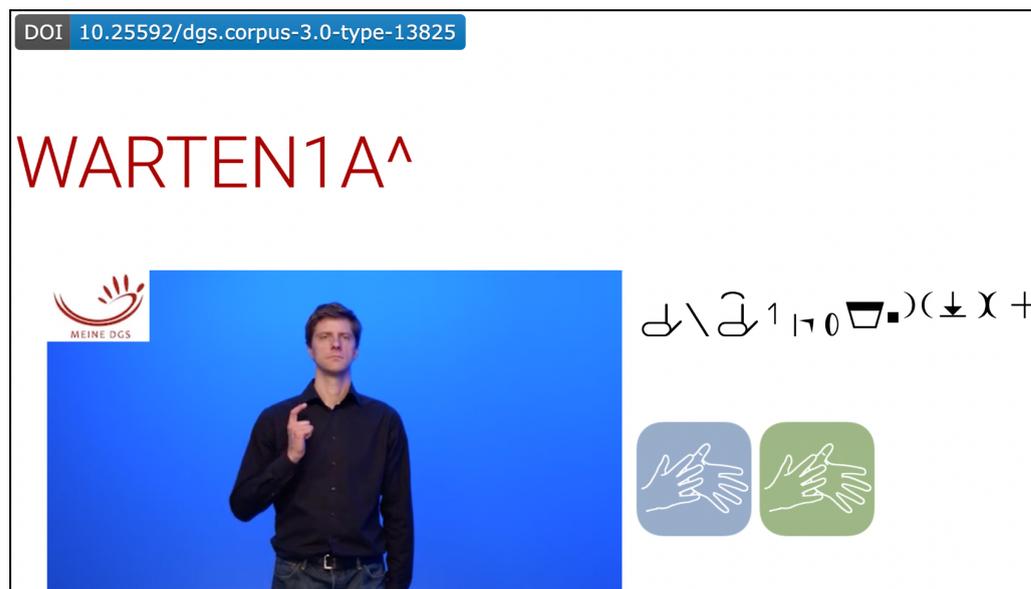
Aqui vemos que, ao contrário do uso de DV, o manual sugere que uma descrição do gesto seja dada. Esta orientação torna a transcrição mais lenta. Entendemos que a mesma lógica aplicada à exclusão da descrição do uso de DV, qual seja o fato de que uma descrição já implicaria numa análise mais específica, o que não é o objetivo da glosagem, deveria ser também aplicada ao uso do “E” para gestos(emblemas), facilitando assim o trabalho de descrição. Os gestos(emblemas) que devem ser glosados como os sinais estáveis devem constar numa lista no manual e/ou no *SignBank*, para que os surdos quais sinais são mais estáveis.

Por fim, para o uso de boia, sugerimos que cada mão acompanhe um sinal, como já é de se esperar, mas que haja um novo código para a mão que está como boia. Os sinais da mão principal que estiverem em relação a mão da boia devem ser glosados com IX, DV, E ou como sinal estável, a depender do contexto, sem perda de informação, já que a mão da boia conterà o código.

#### 5.2.3.2 Caminhos para os problemas de variação fonológica

Para propor um caminho para solucionar este problema, utilizaremos como referência o *corpus* produzido na Alemanha para a língua de sinais alemã (*DGS*). O *corpus* está disponibilizado no site <https://www.sign-lang.uni-hamburg.de/dgs-korpus/index.php/welcome.html> e foi produzido pela Universidade de Hamburgo e conta com vídeos de produções em *DGS* de surdos de toda a Alemanha. Para variações fonológicas, os transcritores acrescentam letras após a transcrição da palavra em alemão. Como já vimos, no *SignBank* da Libras, há a possibilidade de acrescentar números após a palavra em casos de variação lexical, mas não há alternativas para variações fonológicas. No *corpus* da Alemanha, tanto as variações lexicais, quanto fonológicas são contempladas. Vejamos o exemplo a seguir para *WARTEN*, que significa *esperar* em alemão:

Figura 51 - WARTEN1A



Fonte: *DGS Horpus*. Imagem disponível em: [https://www.sign-lang.uni-hamburg.de/meinedgs/types/type13825\\_de.html](https://www.sign-lang.uni-hamburg.de/meinedgs/types/type13825_de.html). Acesso em: 29 de junho de 2022.

Figura 52 - WARTEN1C



Fonte: *DGS Korpus*. imagem disponível em: [https://www.sign-lang.uni-hamburg.de/meinedgs/types/type15630\\_de.html](https://www.sign-lang.uni-hamburg.de/meinedgs/types/type15630_de.html). Acesso em: 29 de junho de 2022.

As figuras mostram que há uma diferença fonológica no uso do sinal, pois na primeira variante (WARTEN1A) o sinal é realizado com a seleção apenas dos dedos polegar

e médio e na segunda variante (WARTEN1C), o sinal é realizado com a seleção do indicador, médio e polegar.

Esta solução poderá dar conta de quaisquer variações fonológicas, é importante estar atento se há de fato uma variação fonológica, ou seja, as duas variantes em questão são estáveis na comunidade surda, ou se trata de somente uma mudança pragmática pontual.

Outro recurso importante que encontramos por acaso no *SignBank* alemão e que nos chamou bastante atenção foi o uso do outro sistema de escrita para descrever o sinal no *SignBank*. Esta alternativa é importante para o registro exato do sinal e pode servir para estudos tanto sobre o *SignWriting*, quanto sobre variação fonológica e lexical, e ainda para pesquisas sobre aprendizagem de *SignWriting* por crianças surdas, como um glossário.

### 5.2.3.3 Caminhos para os problemas de variação lexical

Já há, para variação lexical, a possibilidade de utilizar números para cada variante. No entanto, há uma recomendação de que não sejam alimentadas no *SignBank* muitas variantes, pois isto poderia descaracterizar o perfil padronizador da glosagem. Quanto a isso, a sugestão é que haja, no *SignBank*, um filtro regional que limite a busca por região. Desta forma, as variantes seriam vistas de maneira separada por região, facilitando a busca e futuras pesquisas sobre o léxico.

Um sinal como CONCURSO poderia ser visto no *SignBank* na região sul com duas variantes e na região nordeste com uma ou mais variantes, a depender de quantas variantes sejam encontradas no *corpus*. Não levar estas variantes regionais em conta dificultaria muito o trabalho dos transcritores, pois sabemos que as variações lexicais são muito comuns de região para região.

Além disso, como vimos, muitos sinais não foram encontrados no *SignBank* pela palavra em português, mas pela configuração de mão. Esta estratégia já é prevista (ver Quadros (2016)), mas foi esquecida pelos transcritores. Ainda que a glosa encontrada apresente uma variação com semântica diferente, sendo o mesmo sinal, ele poderá ser encontrado em pesquisas futuras.

#### 5.2.3.4 Caminhos para os problemas de limitação do *SignBank*

Como já discutido, este problema é inevitável, devido à quantidade de itens lexicais que existem na Libras. Portanto, é necessário continuar o trabalho de alimentação de novos sinais no *SignBank*. O manual apresenta a seguinte solução para sinais não conhecidos:

**Tabela 8** - Orientação para sinais não reconhecidos do Manual 2019

Sinal que não é reconhecido.	Cada sinal não reconhecido no enunciado recebe a glosa XXX (pode haver mais de um).	QUERER XXX POR-FAVOR
------------------------------	---	----------------------

Fonte: Manual de transcrição do *Corpus LIBRAS* Inventário Nacional (versão 2019).

Esta estratégia foi mencionada pelos transcritores e está presente em toda esta análise, no entanto, é importante que não só sinais não reconhecidos recebam esta marcação, mas sinais conhecidos que não estão no *SignBank*, ou que apresentam variantes diferentes. Cada dúvida marcada com XXX deve então ser discutida periodicamente em reuniões para validar glosas para estes sinais. Se faz necessário também que seja feita uma transcrição para *SignWriting* de sinais que não constam no *SignBank*, principalmente por conta das variações fonológicas, para que, após a validação do grupo, os sinais sejam gravados e adicionados ao *SignBank*.

Como já dito, estes caminhos não foram testados e representam apenas o início de um percurso para tornar o trabalho de transcrição mais produtivo e preciso. O quadro a seguir apresenta um resumo dos caminhos apresentados nesta seção.

**Tabela 9** - Proposta de solução para problemas de transcrição

PROBLEMAS DE TRANSCRIÇÃO	CAMINHOS PARA SOLUÇÃO
Problemas sociais	Criação de condições para a entrada e permanência de surdos na universidade
	Inserção do aluno surdo em projetos de pesquisa no início da graduação
	Contar com uma quantidade de surdos significativa a partir da disponibilização de bolsas de estudo para que um grupo de trabalho seja formado
Problemas estruturais	Cursos e formações para aprender a utilizar o Manual, o <i>SignBank</i> e o <i>ELAN</i> .
	Formações para desenvolver o conhecimento metalinguístico do surdo a respeito da Libras
	Formações internas para atualização de novos colaboradores por colaboradores mais antigos
	Intercâmbio entre projetos de outros estados para discutir dúvidas e estratégias metodológicas, e para validar novas glosas e sinais
Problemas metodológicos de efeito de modalidade	Distinção clara dos conceitos de Descrição Visual (DV), Gesto (emblema “E”) e sinais estáveis; criação de um novo código para boias;
Problemas metodológicos de variação fonológica	A inserção de letras (a,b,c...) à glosas com variação fonológica (e.g. ALÍVIO-1A, ALÍVIO-1B etc.)
Problemas metodológicos de variação lexical	Busca por configuração de mão no <i>SignBank</i> ; criação de filtro para região, separando as variantes lexicais de região para região do Brasil
Problemas metodológicos de limitação do <i>SignBank</i>	Marcar novos sinais com XXX, transcrever para o <i>SignWriting</i> , discutir os novos sinais e novas glosas em reunião e inserir os novos sinais no <i>SignBank</i>

Fonte: Elaborado pelo autor.

Vale frisar que muito deste percurso já foi pensado e discutido em reuniões com pesquisadores da área de línguas de sinais no mundo, mas especialmente no Brasil, a partir do projeto *Corpus* de Libras muitos documentos sobre restrições do manual de transcrição e novas glosas foram debatidos. O que fizemos aqui foi um levantamento dos problemas a partir da perspectiva do transcritor surdo. Da mesma forma, os caminhos apresentados são

sugestões que partem da visão do transcritor surdo. Combinar estas possíveis soluções pode ser a chave para chegarmos a um processo de transcrição menos desgastante e mais produtivo. É perceptível que muitas estratégias e soluções implicam que outras estratégias também sejam aplicadas para que, conjuntamente, elas possam gerar um resultado efetivo.

Não esperamos com isso que o trabalho de transcrição se torne fácil e rápido, pois isto é impossível. Estes caminhos podem fazer com que os surdos transcritores se sintam mais seguros para trabalhar, pois notamos que a maior parte dos problemas surge da insegurança por não saber qual a melhor tomada de decisão.

## VI. CAPÍTULO - Considerações finais

Após completarmos este percurso de análise das entrevistas, fazer um levantamento dos problemas e de transcrição e propor caminhos para solucionar os problemas levantados, entendemos que os nossos objetivos foram alcançados, pois fizemos um extenso levantamento dos problemas de transcrição e propusemos soluções e estratégias para estes problemas. Subdividimos os problemas de transcrição em 3 tipos: problemas sociais (individuais, locais e gerais), problemas estruturais e problemas metodológicos (de efeitos de modalidade, de variação fonológica, de variação lexical e de limitação do *SignBank*).

Para lidar com estes problemas, listamos as seguintes estratégias e soluções: a criação de condições para a entrada e permanência de surdos na universidade; a inserção do aluno surdo em projetos de pesquisa no início da graduação; contar com uma quantidade de surdos significativa a partir da disponibilização de bolsas de estudo para que um grupo de trabalho seja formado; a realização de cursos e formações para aprender a utilizar o Manual, o *SignBank* e o *ELAN*; formações para desenvolver o conhecimento metalinguístico do surdo a respeito da Libras; formações internas para atualização de novos colaboradores por colaboradores mais antigos; intercâmbio entre projetos de outros estados para discutir dúvidas e estratégias metodológicas, e para validar novas glosas e sinais; estabelecer uma distinção clara dos conceitos de Descrição Visual (DV), Gesto (emblema ‘E’) e sinais estáveis; a criação de um novo código para bóias; a inserção de letras (a,b,c...) às glosas com variação fonológica (e.g. ALÍVIO-1A, ALÍVIO-1B etc.); a busca por configuração de mão no *SignBank*; criação de filtro para região, separando as variantes lexicais de região para região do Brasil; marcar novos sinais com XXX, transcrever para o *SignWriting*, discutir os novos sinais e novas glosas em reunião e inserir os novos sinais no *SignBank* com a glosa e a versão em *SignWriting*.

Como possibilidades para novas pesquisas, sugerimos que testes sejam feitos a partir deste caminho, pois é possível que algumas estratégias precisem ser adaptadas levando em conta diferentes realidades, ou novas problemáticas não observadas neste trabalho. Além disso, como o trabalho de tradução para o português e de revisão ainda não foram realizados, seria importante investigar quais questões estão por trás também destes processos para que a construção do *corpus* seja finalmente concluída, e que novas construções de *corpus* possam levar em conta, não só as contribuições deste trabalho para as transcrições, mas de outros trabalhos sobre todas as etapas da criação de um *corpus* de línguas de sinais.

## REFERÊNCIAS

BAGNO, M. (Org.). *Nada na língua é por acaso: por uma pedagogia da variação linguística*. São Paulo: Parábola, 2007.

BARROS, M. E. *Elis – escrita das línguas de sinais: proposta teórica e verificação prática*. 2008. 199 f. Tese (Doutorado em Linguística) – Curso de Pós-Graduação em Linguística, Universidade Federal de Santa Catarina, Florianópolis, 2008.

BRASIL. *Decreto nº 10.502*, de 30 de setembro de 2020. Disponível em: <[http://www.planalto.gov.br/ccivil\\_03/\\_ato2019-2022/2020/decreto/D10502.htm](http://www.planalto.gov.br/ccivil_03/_ato2019-2022/2020/decreto/D10502.htm)>. Acesso em: 11 ago. 2021.

BRASIL. *Decreto nº 5.626*, de 22 de dezembro de 2005. Disponível em: <[http://www.planalto.gov.br/ccivil\\_03/\\_Ato2004-2006/2005/Decreto/D5626.htm](http://www.planalto.gov.br/ccivil_03/_Ato2004-2006/2005/Decreto/D5626.htm)>. Acesso em: 27 jul. 2021.

BRASIL. *Lei nº 10.436*, de 24 de abril de 2002. Disponível em: <[http://www.planalto.gov.br/ccivil\\_03/leis/2002/110436.htm](http://www.planalto.gov.br/ccivil_03/leis/2002/110436.htm)>. Acesso em: 27 jul. 2021.

BRASIL. *Lei nº 14.191*, de 3 de agosto de 2021. Disponível em: <[http://www.planalto.gov.br/ccivil\\_03/\\_ato2019-2022/2021/Lei/L14191.htm](http://www.planalto.gov.br/ccivil_03/_ato2019-2022/2021/Lei/L14191.htm)>. Acesso em: 11 ago. 2021.

BRASIL. Presidência da República. *Decreto nº 7.387*, de 09 de dezembro de 2010. Institui o Inventário Nacional da Diversidade Linguística e dá outras providências. Brasília, 2010. Disponível em: <[http://www.planalto.gov.br/CCIVIL\\_03/\\_Ato2007-2010/2010/Decreto/D7387.htm](http://www.planalto.gov.br/CCIVIL_03/_Ato2007-2010/2010/Decreto/D7387.htm)>. Acesso em: 27 jul. 2021.

CEZARIO, M. M., VOTRE, S.. Sociolinguística. In: Martelotta, M.E. (Org.) et al. *Manual de Linguística*. São Paulo: Contexto, 2016.

COELHO, Izete Lehmakuhl. GORSKI, Edair Maria. SOUZA, Christiane Maria N. MAY, Guilherme Henrique. *Para Conhecer: Sociolinguística*. São Paulo: Contexto, 2015.

IPHAN. *INDL: Pesquisa coleta dados para o Inventário Nacional de Libras*. Disponível em: <<http://portal.iphan.gov.br/indl/>>. Acesso em: 27 jul. 2021.

FENLON, J., SCHEMBRI, A., JOHNSTON, T. E CORMIER, K. *10 0 Documentary and Corpus Approaches to Sign Language Research*. Published, 2015.

GAMA, F. J. DA. *Iconographia dos Signaes dos SurdosMudos*. Rio de Janeiro, 2011. (Série Histórica do Instituto Nacional de Educação de Surdos. v. 1). Disponível em: <<https://drive.google.com/file/d/0B6WyKJSZvdJdb0M2RjhqcFVBOG8/view>>. Acesso em: 27 jul. 2021.

HEITKOETTER, R. P.; XAVIER, A. N. *Descrição e análise de boias de listagem em libras*, 2021. Disponível em: <https://revista.unitins.br/index.php/humanidadeseinovacao/article/view/3234>. Acesso em 27 jun 2022.

KARNOPP, L. (s/d). *Fonética e fonologia. Apostila do curso de Letras-Libras-licenciatura e bacharelado*. Florianópolis: UFSC.

LABOV, William (1972). *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press. [Padrões Sociolinguísticos. Trad.: Marcos Bagno; Marta Scherre e Caroline Cardoso. São Paulo: Parábola, 2008.]

LEITE, T. DE A. E QUADROS, R. M. DE. *Línguas de Sinais do Brasil: Reflexões sobre o seu estatuto de risco e a importância da documentação*. Em Estudos da Língua de Sinais. Volume II. Editora Insular, 2014.

MOLLICA, Maria Cecília. *Introdução à Sociolinguística: o tratamento da variação*. São Paulo: Contexto, 2015.

NONAKA, A. (2004). *The forgotten endangered languages: Lessons on the importance of remembering from Thailand's Ban Khor Sign Languages*. *Language in Society*, v. 33, p. 737-767.

NYST, V. *7 Sign Language Fieldwork*. Published, 2015.

O' KEEFFE, A., McCARTHY, M. *Historical perspective What are corpora and how have they evolved*. *The Routledge Handbook of Corpus Linguistics*, 2010.

OLIVEIRA Jr. M. *Técnicas de recolha de dados orais para documentação linguística*. In: LEITE, C. T. e SILVA, J. B. *Línguas no Brasil: coleta, análise e descrições de dados*. Maceió: EDUFAL, 2013.

OUSHIRO, Livia. *Construção de uma amostra da fala paulistana. Manual de Transcrições no ELAN*. Projeto SP 2010.

QUADROS, R. M. DE.; KARNOPP, Lodenir Becker. *Língua de Sinais brasileira: Estudos linguísticos*. Porto Alegre: ArtMed, 2004.

QUADROS, R. M. DE.; LEITE, T. DE A.; LOHN, J. T.; SCHMITT, D. PEGO, C. *Documentação da Língua Brasileira de Sinais*, 2020.

QUADROS, R. M. DE. A transcrição de textos do *Corpus de Libras*. *Revista Leitura*, Maceió V.1 no 57 – jan/jun p. 8 - 34, 2016.

QUADROS, R. M. DE. *Documentação da Língua Brasileira de Sinais*. In: SEMINÁRIO IBERO-AMERICANO DE DIVERSIDADE LINGUÍSTICA. Anais [...]. Brasília, DF: Iphan, 2016a.

QUADROS, R. M. DE. et al. *Corpus de Libras*. Florianópolis, UFSC, 2017c. Disponível em: <http://corpuslibras.ufsc.br/>. Acesso em: 30 julho. 2021.

QUADROS, Ronice Müller de et al. *Língua brasileira de sinais: patrimônio linguístico brasileiros*. Florianópolis: Editora Garapuvu, 2018.

ROYER, M. *Análise da ordem das palavras nas sentenças em Libras do Corpus da Grande Florianópolis*. Dissertação (Mestrado em Linguística) - UFSC. SC, 2019.

SARDINHA, Tony Berber. *Linguística de Corpus: histórico e problemática*. DELTA, São Paulo, v. 16, n. 2, p. 323-367, 2000.

SINGLETON, J. L., AMBER J. MARTIN, AND GARY MORGAN. *1 Ethics, Deaf-Friendly Research, and Good Practice When Studying Sign Languages*. Published, 2015.

STROBEL, K. L. *As imagens do outro sobre a cultura surda*. Florianópolis: Editora UFSC, 2008.

\_\_\_\_\_. *Surdos: vestígios culturais não registrados na história*. 2008. 176p. Tese (Doutorado em Educação), Universidade Federal de Santa Catarina, Florianópolis, 2008.

UFSC. Libras. *SignBank da Libras*, 2020. Disponível em: <<http://signbank.libras.ufsc.br/>>. Acesso em: 02 ago. 2022.

UNESCO, Disponível em: <<http://envolverde.cartacapital.com.br/brasil-t-em-190-idiomas-sob-risco-de-extincao-segundo-unesco/>> Acesso em 27 jul. 2021.

VAN HERREWEGHE, M., E VERMEERBERGEN, M. (2012). *Data collection. Sign language: an international handbook*.

VILHALVA, S. *Índios Surdos: Mapeamento das Línguas de Sinais do Mato Grosso do Sul*. Petrópolis, RJ: Arara Azul, 2012.