

DOCUMENTAÇÃO DE LÍNGUAS DE SINAIS

DOCUMENTACIÓN DE LAS LENGUAS DE SEÑAS

DOCUMENTATION OF SIGN LANGUAGES

Ronice Müller de Quadros*

Universidade Federal de Santa Catarina

Christian Rathmann**

Humboldt-Universität zu Berlin

Johanna Mesch***

Stockholm University

Jair Barbosa da Silva****

Universidade Federal de Alagoas

*... research in the Deaf community should be by Deaf, for Deaf, and with Deaf people.
... pesquisa na comunidade Surda deveria ser feita pelo Surdos, para o Surdo e com pessoas surdas.
Raychelle Harris, Heide M. Holmes and Donna M. Mertens (tradução dos autores)*

RESUMO: Este artigo apresenta uma discussão sobre a documentação de línguas de sinais, considerando motivações acadêmicas e sociais, assim como indicações de materiais que as comunidades surdas consideram relevantes integrarem tais documentações. Primeiramente, iremos apresentar uma discussão de ordem mais filosófica acerca da documentação linguística das línguas de sinais e, posteriormente, abordaremos as questões de ordem técnica e ética relativas à elaboração de *corpora* nessas línguas. A primeira parte do artigo, abordará três questões básicas: (1) o que constitui a documentação de línguas de sinais; (2) para quê documentar as línguas de sinais; (3) para quem é necessário documentar as línguas de sinais. A segunda parte do artigo trará práticas de documentação de línguas de sinais em diferentes países como referência de aproximação e valorização das comunidades surdas. Serão apresentadas práticas de documentação, em especial, a constituição de *corpora* de línguas de sinais enquanto língua das comunidades surdas; registro histórico das línguas de sinais; produções literárias em línguas de sinais; referências para o ensino de línguas de sinais; registro de produções de surdos bilíngues unimodais; registro de produções de bilíngues bimodais. Estas questões

* Professora e pesquisadora da Universidade Federal de Santa Catarina, no Departamento de Libras e pesquisadora do CNPQ. E-mail: ronice.quadros@ufsc.br.

** Professor e pesquisador da Humboldt-Universität zu Berlin. E-mail: christian.rathmann@hu-berlin.de.

*** Professora e pesquisadora da Stockholm University, na Suécia. E-mail: johanna.mesch@ling.su.se.

**** Professor e pesquisador da Universidade Federal de Alagoas. E-mail: jair.silva@fale.ufal.br.

estão permeadas por aspectos éticos, uma vez que a documentação de línguas de sinais envolve as comunidades surdas que usam suas respectivas línguas de sinais. Ao final, faremos uma reflexão de como estas práticas de documentação se voltam para os interesses acadêmicos e sociais das comunidades surdas.

PALAVRAS-CHAVE: Documentação de línguas. Línguas de sinais. Comunidades surdas.

RESUMEN: Este artículo presenta una discusión general sobre la documentación de las lenguas de señas, considerando sus motivaciones académicas y sociales, así como algunas recomendaciones de materiales que las comunidades sordas consideran relevantes para integrar dicha documentación. En primer lugar, presentaremos una discusión más filosófica sobre la documentación lingüística de las lenguas de señas y, posteriormente, abordaremos cuestiones técnicas y éticas relacionadas con la elaboración de *corpus* en estas lenguas. La primera parte del artículo, abordará tres cuestiones básicas: (1) ¿qué constituye la documentación en lenguas de señas?; (2) ¿para qué documentar las lenguas de señas?; (3) ¿para quién es necesario documentar las lenguas de señas? La segunda parte del artículo traerá prácticas de documentación de lenguas de señas como referencia para acercarse y valorar a las comunidades sordas. Se presentarán prácticas de documentación, en particular, la constitución de *corpus* de la lengua de señas como la lengua de las comunidades sordas, registro histórico de las lenguas de señas, producciones literarias en lenguas de señas, referencias para la enseñanza de lenguas de señas, registro de producciones de sordos bilingües unimodales y registro de producciones de bilingües bimodales. Estos temas están atravesados por aspectos éticos, ya que la documentación de lenguas de señas involucra a las comunidades sordas que utilizan sus respectivas lenguas de señas. Al final, reflexionaremos sobre cómo estas prácticas de documentación abordan los intereses académicos y sociales de las comunidades sordas.

PALABRAS CLAVE: Documentación de lenguas. Las lenguas de señas. Comunidades sordas.

ABSTRACT: This article presents a discussion about documentation of sign language, with special attention to the academic and social aspects, and a few recommendations as to what kinds of materials are considered relevant by Deaf communities in sign language documentation. First, we take a more philosophical approach in discussing the linguistic documentation of sign languages, after which we address technical and ethical aspects of compiling linguistic *corpora* of signed languages. As such, the first part of this article focuses on three basic questions: (1) what exactly constitutes documentation of sign languages; (2) why document sign languages at all; and (3) on whose behalf is it necessary to document sign languages. The second part of this article explores a number of exemplary practices in sign language documentation as executed in various countries that are highly valued and may serve as points of reference for diverse Deaf communities. Documentation practices will be presented, in particular, the creation of *corpora* to represent sign languages as they are used by Deaf communities; historical records of sign languages; literary production in sign languages; reference materials for teaching sign languages; records of bilingual-unimodal Deaf signers; and records of bilingual-bimodal signers. These issues carry with them numerous ethical implications, as the documentation of sign languages inevitably involves the Deaf communities of each respective language. As such, we close the article with a reflection on how these documentation practices relate to the academic and social interests of Deaf communities involved in such research.

KEYWORDS: Documentation of languages. Sign languages. Deaf communities.

1 INTRODUÇÃO

A palavra documentação remete à reunião de documentos que são fontes de informação registradas e devidamente organizadas para subsidiar alguma coisa. No caso específico da documentação de línguas, o termo envolve a reunião de dados para registrar línguas com o objetivo de servir de referência para diferentes fins: pesquisa linguística, históricos, educacionais, culturais, literários, dentre outros. Esta definição é bastante simplista, mas dá uma ideia geral do que iremos discutir neste artigo: a documentação das línguas de sinais.

As primeiras documentações de línguas foram constituídas de registro de termos para a composição de dicionários. Segundo McCarthy e O'Keeffe (2010), um dos primeiros registros de línguas foi feito para compor glossários da bíblia sagrada com o objetivo de especificar palavras contidas neste material altamente difundido na antiguidade. Posteriormente, os estudiosos começaram a registrar termos de obras literárias, como, por exemplo, Shakespeare. A partir disso, a documentação de línguas começou a ser

expandida e foram constituídos os *corpora* de línguas, o que teve como consequência o desenvolvimento de técnicas e de recursos metodológicos (sobretudo com o uso e propagação dos programas computacionais) constitutivos da Linguística de Corpus.

A documentação de línguas passou a integrar os *corpora* de línguas que se apresentam características específicas para fins voltados para a pesquisa linguística. Veja que "documentação" inclui "*corpus*", pois as línguas podem ser documentadas para diferentes fins com níveis de sistematização bastante diferentes. No caso específico de *corpora* de línguas, a documentação torna-se altamente especializada, pois passa a atender uma série de princípios mais técnicos para servir de registro para fins de pesquisa linguística contendo metadados que facilitem o processo de investigação.

Mccarthy e O'Keeffe (2010) apresentam um histórico sobre a ideia de *corpus* linguístico. Segundo os autores, a proposta de *corpus* começou a tomar forma na década de 60, quando os estudos do campo de lexicografia começaram a se desenvolver e a tecnologia começou a viabilizar formas de registro diferenciadas e em quantidades maiores. O primeiro *corpus* computadorizado foi o *Corpus* de Brown, compilado em 1960 na Brown University, contendo um milhão de palavras do inglês de textos literários. Em 1970, houve a consolidação da linguística de *corpus* com várias línguas e com diferentes propostas. Em 1990, a invenção do scanner foi uma revolução para a documentação de textos escritos. Na virada do milênio, os textos passam a ser digitalizados de forma ilimitada, viabilizando a composição de *corpora* de línguas com dimensões gigantescas. Atualmente, o Google é, por si só, uma forma de documentação de uma infinidade de textos que chega a permitir a computação de dados para fomentar a tradução automática. Diante dessa explosão de documentações, o problema do linguista mudou de foco: de acessar grandes quantidades de dados para elaborar metodologias confiáveis para descrever e dar conta das evidências linguísticas permitidas pelos dados organizados nos *corpora*.

A documentação de línguas para fins linguísticos começou a focar nos princípios da constituição de *corpora*. Segundo Rappen (2010), dependendo o tipo de questão a ser pesquisada, a construção do *corpus* pode ser uma tarefa razoavelmente eficiente e restrita, ou bastante ampla e demorada. Ter uma questão claramente definida é o primeiro passo na construção do *corpus*. Outro princípio básico é de que um *corpus* de língua deve ser representativo. O tamanho do *corpus* não é único. Não há um número específico de palavras e dados. A questão sobre o tamanho do *corpus* pode ser resolvida por dois fatores: (a) a representatividade (precisamos coletar textos e palavras suficientes para representar com exatidão a língua pesquisada); e (b) a praticidade (o tempo deve ser o mais restrito possível), ou seja, dados suficientes para representar o que está sendo estudado, sem exceder o que seja necessário de acordo com o tempo disponível para a realização de um determinado estudo. Rappen (2010) menciona que depois de a questão de pesquisa estar definida, a construção do *corpus* pode começar. A tarefa seguinte será identificar os textos e desenvolver um plano para a coleta. Quando a coleta for feita com pessoas e/ou instituições é essencial ter o consentimento das partes envolvidas, ter permissão (termo de consentimento assinado pelos participantes), o que implica uma questão ética a ser observada quando da elaboração de um *corpus* linguístico.

Rappen (2010) levanta um conjunto de questões que precisam ser respondidas antes de se criar um *corpus*, independentemente de ele estar representando a língua falada/sinalizada ou escrita:

- I. O que constitui um texto?
- II. Como os arquivos devem ser nomeados?
- III. Que informações serão incluídas em cada arquivo?
- IV. Como os textos serão armazenados (formato do arquivo)?
- V. Como os textos serão catalogados?

Estas questões gerais são importantes e evidenciam a necessidade de pré-definir um conjunto de procedimentos para realizar uma documentação que seja apropriada, especialmente quando os registros servirão de fonte de pesquisa. Por exemplo, um texto pode ser o registro de textos (orais, sinalizados, escritos etc.) diferentes que serão delimitados pelos objetivos propostos. No caso dos primeiros *corpora* estabelecidos, os textos envolveram a bíblia sagrada e textos literários. A forma de organização dos registros envolve definir se será escrito, áudio ou vídeo, por exemplo. Quando coletados, como estes dados serão categorizados? Uma possibilidade é definir pastas por dia de coleta, por tipo de texto, pelo nome do participante, por exemplo. O nome do arquivo pode

registrar claramente seu conteúdo, mostrando aspectos relevantes do texto para análise. Isso permite ao usuário classificar e agrupar em subcategorias de arquivos, ou criar sub *corpora* mais facilmente. Metadados precisam ser associados aos dados documentados para ser usados para fins de contextualização do dado (informações sobre os participantes, data da coleta, tipo de texto, formato do arquivo). A etapa de arquivamento dos registros também é muito importante, pois facilitará a sua localização, além do armazenamento propriamente dito. O problema do armazenamento está relacionado com o espaço, pois o tamanho do *corpus* e o tipo de material pode demandar espaços gigantescos de armazenamento. Isso tudo precisa ser considerado ao se documentar uma língua, sobretudo levando-se em consideração o formato da documentação. Por fim, a catalogação vai envolver também a anotação dos dados por meio de transcrições e estabelecimento de categorias de análise. Para isso, torna-se necessário registrar detalhadamente como será feita a transcrição, que ferramentas/softwarewares serão usados, quais os níveis de anotação serão registrados. Tudo isso precisa ser feito para viabilizar um sistema de busca eficiente.

No caso específico da documentação das línguas de sinais, a tecnologia propiciou a documentação no formato de vídeo viabilizando o seu registro. No entanto, há uma grande complexidade envolvida no registro de textos em línguas de sinais, uma vez que estas línguas são produzidas pelo corpo (mãos, face, tronco corporal) de forma tridimensional. Isso requer o registro em vídeos considerando diferentes ângulos da sinalização. A tecnologia tem avançado muito e as câmeras 3D já começaram a ser comercializadas, mas de modo geral, os registros têm sido feitos com várias câmeras combinadas para registrar diferentes perspectivas do corpo a fim de captar esta tridimensionalidade intrínseca destas línguas. A exemplo, apresentamos o uso de quatro câmeras simultâneas que registram a interação entre duas pessoas utilizando a língua de sinais para tentar captar o uso do corpo e do espaço que integram o texto das línguas de sinais.



Figura 1: Utilização de quatro câmeras para captar diferentes perspectivas da produção em sinais de duas pessoas conversando

Fonte: *Corpus* de Libras da UFSC

A complexidade envolvida no registro línguas de sinais envolve também o processo de anotação linguística dos sinais, uma vez que estas línguas ainda não contam com um sistema escrito difundido. Há alguns sistemas de escritas que têm sido usados e que requerem softwares independentes que permitem o registro dos sinais. No entanto, de modo geral, a maioria dos linguistas utilizam sistemas de anotação com base em sistemas de glosas que apresentam muitas limitações em face da tridimensionalidade das línguas de sinais e, sobretudo, das marcações não manuais com valor linguístico presentes nessas línguas. Diante disso, os *corpora* de línguas de sinais requerem ferramentas que permitam a visualização da língua de sinais sempre que os dados forem analisados. Um sistema de anotação amplamente usado é o *Eudico Annotator ELAN* (2020).

2 PARA QUÊ DOCUMENTAR AS LÍNGUAS DE SINAIS?

A documentação de uma língua pode atender a propósitos diversos, os quais vão desde razões de ordem técnica, como compilar as palavras que compreendem uma língua em forma de dicionários e glossários, por exemplo, até atitudes de política linguística, como salvar línguas em risco de extinção, fato frequente ao redor do mundo por razões políticas, econômicas, históricas, a exemplo do que ocorre com as línguas indígenas. Neste sentido, documentar uma língua é a garantia, por vezes, de revitalizar essa língua, ou, por outras, apenas manter um registro histórico dessa língua. O planejamento linguístico, as políticas de valorização e de manutenção do status linguístico de uma dada língua requerem documentação dessa língua. Para além disso, a descrição linguística se faz imperativa e, para tanto, os dados documentados são objeto de análise. Do ponto de vista linguístico mais amplo, um *corpus* linguístico pode subsidiar o pesquisador (linguista) para lidar com vasto espectro da linguagem humana a partir de diferentes práticas linguísticas (usos da língua).

Considerando que a linguagem humana se manifesta por meio de textos, os quais têm uma dada estrutura gramatical, mas também veiculam discursos, os *corpora* em linguística são sempre elaborados atendendo a certos princípios norteadores do ponto de vista técnico, bem como aos propósitos da pesquisa. Para Johnston (2009) a Linguística de Corpus tem como princípio a suposição de que uma grande quantidade de dados anotados e processados revelam certos padrões linguísticos sobre o uso e estrutura das línguas incapazes de serem alcançados pela introspecção dos pesquisadores, mesmo dos mais especializados. Assim, argumenta o autor, em Linguística de Corpus, "[...] quantitative analysis goes hand in hand with qualitative analysis!" (JOHNSTON, 2009, p. 18). Johnston (2014) também aborda sobre a importância de adicionar valores e extrair valores das descrições linguísticas por meio de anotações estratégicas que compreendem *corpora* de línguas de sinais.

Aliados aos aspectos linguísticos, tecnológicos e metodológicos constitutivos dos *corpora*, os aspectos éticos são tão relevantes quanto os demais. Questões éticas devem permear as razões que justificam a documentação de línguas, pois percorrem todo o planejamento que precisa envolver princípios éticos que ampliem a visão dos pesquisadores. Estas razões têm justificado a documentação das línguas de sinais.

Alguns princípios éticos mais gerais precisam ser considerados na documentação de línguas de sinais:

- a) O respeito pelos valores linguísticos, culturais, sociais, morais, religiosos, assim como costumes das comunidades surdas envolvidas.
- b) A garantia de benefícios às comunidades surdas envolvidas na documentação que se estendem para além do período da realização da coleta de dados.
- c) A consideração à diversidade envolvida nas comunidades surdas precisa ser levada em conta na consolidação da documentação das línguas.
- d) O fato de que pesquisa com surdos precisa ser feita por surdos, mesmo que os ouvintes estejam envolvidos (HARRIS; HOLMES; MERTENS, 2019).

Estes princípios precisam guiar o porquê da documentação de uma dada língua de sinais. Os surdos relatam que querem se enxergar na documentação, querem ser devidamente identificados enquanto autores do seu próprio texto, querem ver os vídeos que compõem o todo e, também, querem perceber nos materiais a herança surda.

Segundo Harris, Holmes e Mertens (2009), entre as várias perguntas que devem ser consideradas ao se pensar no para quê fazer a documentação de uma língua de sinais, as seguintes devem provocar reflexões:

- a) Quais são os princípios culturais mais apropriados para este grupo?
- b) O que incluiremos nas diretrizes para indicar o respeito e consideração a sua cultura?
- c) Como apresentamos a importância dos princípios culturais que regem esta população?

¹ A análise quantitativa anda de mãos dadas com a análise qualitativa.

d) Como os pesquisadores que estão conduzindo pesquisas em comunidades culturais complexas podem incorporar as vozes de seus membros quando diante de questões éticas e metodológicas?

Estas questões aplicadas às comunidades surdas precisam ser consideradas cuidadosamente pelos pesquisadores. No caso dos surdos, a complexidade cultural é permeada pelas relações entre os surdos e com os ouvintes. Um pesquisador ouvinte que pretende desenvolver uma pesquisa com as comunidades surdas precisará ter muita sensibilidade, respeito e ser justo no propósito de responder a estas questões juntamente com membros das comunidades pesquisadas, com a presença de pessoas surdas. Para tanto, como abordado ainda por Harris, Holmes e Mertens (2009), estes pesquisadores precisam desenvolver uma competência cultural, ou seja, estarem abertos, atentos, dispostos a aprender sobre esta comunidade, compreendendo e apreciando o contexto cultural no qual a pesquisa será implementada com a perspectiva de transformar as intervenções de tal forma que se tornem realmente legítimas aos participantes. Ainda e não menos importante, os pesquisadores precisam realmente ser sensíveis ao que os surdos pensam e sentem sobre o mundo e suas relações pessoais, reconhecendo de fato esses sentimentos e pensamentos porque reconhecem que eles merecem esse reconhecimento. Só assim é possível propor uma documentação da língua de sinais dessas comunidades.

Para além dos aspectos culturais relativos ao respeito às comunidades surdas elencados por Harris, Holmes e Mertens (2009), é importante ressaltar que a elaboração de *corpora* em línguas de sinais necessariamente envolve a exposição do surdo, uma vez que a expressão linguística nessas línguas se dá por meio de corpo (movimentos de mãos e expressões faciais) e espaço da sinalização. Assim, diferentemente do que ocorre com línguas orais ou com a escrita, a codificação dos dados não permite ocultar o participante da pesquisa. Dito de outra forma, a imagem do informante surdo estará sempre à mostra nos vídeos que constituirão os *corpora*, portanto, há de se estabelecer boas práticas de pesquisa, balizadas pela ética, de modo a não se violar aspectos culturais, identitários e históricos dos informantes surdos, além de lhes permitir compensação - no caso do Brasil, essa compensação não pode ser de ordem financeira, mas em termos de retorno à comunidade surda dos dados, dos produtos gerados a partir da documentação.

A documentação das línguas de sinais, portanto, é um trabalho científico de extrema relevância, de natureza interdisciplinar e plural, com vistas ao registro dessas línguas, sua disseminação, vitalização, descrição e guarda como patrimônio imaterial de um povo, o surdo.

3 A QUEM SERVE A DOCUMENTAÇÃO DE LÍNGUAS DE SINAIS?

A pergunta "a quem serve a documentação de línguas de sinais?" é fundamental, pois envolve também aspectos de ordem ética. As comunidades surdas têm participado da constituição de vários *corpora* de línguas de sinais. Historicamente, os surdos são convidados a serem informantes de pesquisadores de línguas de sinais e nos deparamos com relatos de surdos que manifestam sua sensação de serem explorados.

No Brasil, vários surdos que foram consultados quanto a sua língua de sinais nas décadas de 80 e 90 relatam que não participaram efetivamente das pesquisas e que se sentiam explorados. O que significa este sentimento de "exploração"? De modo geral, os relatos sobre isso são muito claros: a sensação de servir de informante fornecendo dados e discutindo sobre eles sem participar efetivamente da pesquisa e sem entender claramente para que serve a pesquisa foi uma realidade. Estes surdos não foram autores destas pesquisas ou não receberam efetivamente nenhum retorno de seus resultados.

Por trás desta insatisfação está claramente colocada uma necessidade das comunidades surdas de serem autores dos estudos da própria língua. Quem pesquisa uma língua? Normalmente os próprios falantes da língua são os protagonistas das pesquisas com suas respectivas línguas. Até é possível contarmos com pesquisadores que não sejam falantes de uma língua incluírem em seus estudos aspectos linguísticos de análise de outra língua, mas de modo geral, os alemães pesquisam a língua alemã, os ingleses e os americanos pesquisam o inglês, os franceses pesquisam a língua francesa e assim por diante. Quem são os pesquisadores de línguas de sinais? Na grande maioria não são os próprios surdos e isso provoca uma série de reações entre os surdos, dentre elas, o sentimento de serem explorados.

A quem serve, então, a documentação das línguas de sinais? A resposta não deveria ser de que serve aos linguistas (que, na grande maioria, ainda nem são surdos!), mas sim devem servir aos interesses das comunidades surdas, aos próprios surdos para que assistam e participem da valorização de suas línguas, por meio de seus registros e dos resultados das pesquisas decorrentes de sua documentação. Há uma mudança no posicionamento dos pesquisadores em relação à realização de pesquisas com surdos, no sentido de garantir um trabalho colaborativo (HOCHGESANG, 2015), o que implica em decidir para quê a pesquisa servirá.

O discurso dos pesquisadores não basta na apresentação de um termo de consentimento no qual os participantes podem aceitar ou recusar a sua participação na documentação de sua própria língua. Os pesquisadores precisam estar comprometidos com as comunidades surdas envolvidas, ou seja, precisam conhecer profundamente os valores que constituem a relação com sua língua a partir do convívio com as comunidades surdas. O primeiro passo, portanto, é ser profundamente fluente na língua de sinais destas comunidades. A partir da língua, os pesquisadores precisam conhecer o que rege as relações interpessoais de seus participantes na relação entre os próprios surdos e na relação com os ouvintes. O pesquisador enquanto surdo ou ouvinte precisará estar muito consciente destas relações para poder discutir sobre elas e planejar a documentação da língua.

Singleton *et al.* (2014) investigaram o quanto os participantes surdos efetivamente dão a sua aprovação ao assinarem o termo de consentimento. Eles entrevistaram surdos que participaram de pesquisas e identificaram que muitos deles sentem-se constrangidos ou não confortáveis diante dos pesquisadores que pode ser decorrente tanto da falta de sensibilidade cultural, como das limitações na comunicação. Um exemplo dado pelos autores é da apresentação do termo de consentimento na forma escrita, sem uma versão na língua de sinais. Outro fator identificado é que devido às limitações na comunicação, muitas vezes o que o participante apresentou não era exatamente o que havia sido solicitado levando a resultados que não representam o que realmente poderia ser obtido. Singleton *et al.* (2014), então, apresentaram alguns pontos a serem considerados no desenvolvimento de pesquisas: (1) incorporação da consciência cultural e linguística; (2) criação de práticas acessíveis com termos em língua de sinais e processos de desenvolvimento da pesquisa na língua de sinais e na língua escrita organizados de forma acessível; (3) o retorno dos resultados das pesquisas aos participantes, especialmente aos que manifestarem interesse em conhecer os resultados da pesquisa; (4) reestruturar modelos colaborativos envolvendo surdos e ouvintes de forma transparente. Essas considerações éticas são importantes nas propostas de documentação de línguas de sinais.

O porquê e o para quem precisam ter sempre como objetivo maior da documentação servir aos surdos e suas comunidades. Os direitos e a integridade dos surdos devem ser a base das propostas de documentação de línguas de sinais. Para isso, é importante situar temporalmente cada parte integrante da documentação, elucidando o seu contexto e a sua função social no momento da sua constituição. Os materiais a integrem a documentação podem passar por gerações e, por isso, precisam ser devidamente contextualizados, pois envolvem participantes de um determinado tempo e de um contexto específico. Outro aspecto importante é a questão da comunicação, pois várias pesquisas não comunicam devidamente por causa das limitações dos pesquisadores em relação à língua de sinais usada pela comunidade envolvida. É fundamental a presença de surdos pesquisadores que pertençam à comunidade. Isso garante o respeito à integridade daqueles que aceitaram participar de um determinado projeto que integrará a documentação. Ao pesquisador compete ser responsável pela promoção das pessoas surdas e de suas comunidades por meio da documentação. Assim, garantimos a integridade de quem participa na consolidação da documentação de sua língua.

Independente de quem participa na documentação das línguas de sinais, precisamos propiciar que os próprios surdos tornem-se autores da documentação de suas línguas de sinais (HOCHGESANG, 2015). Nos processos de documentação de línguas de sinais, muitas vezes os pesquisadores ouvintes ficam constrangidos por não serem surdos e, às vezes, não usuários de língua de sinais. É importante discutirmos sobre isso, pois os autores surdos sempre reforçam a importância de surdos estarem decidindo para quê e para quem servem as pesquisas, especialmente, as documentações de suas línguas de sinais e sempre citam que contam com pesquisadores ouvintes que são sensíveis aos surdos, são usuários da língua de sinais e são verdadeiros parceiros nas proposições de pesquisa. Assim, os pesquisadores ouvintes não precisam se sentir constrangidos se efetivamente estiverem trabalhando juntamente com pesquisadores surdos e, muitas vezes ainda, motivando os surdos a assumirem estas posições. Neste sentido, as pesquisas colaborativas contando com surdos e ouvintes podem ser pesquisas que partem do olhar dos próprios surdos para a constituição de diferentes tipos de documentação de línguas de sinais. Nos termos de Singleton, Martin e Morgan (2014, p. 08) "When hearing

researchers work closely with Deaf researchers, the resulting collaboration can bring positive rewards; but it does not come without preparation and conscientious effort²”.

4 DOCUMENTAÇÃO DE LÍNGUAS DE SINAIS EM DIFERENTES PAÍSES

Nesta seção, apresentamos alguns projetos de documentação em diferentes línguas de sinais que atentaram a importância do envolvimento das comunidades surdas locais no estabelecimento da documentação de suas línguas de sinais.

Hochgesang (2015) apresenta um projeto de documentação da língua de sinais do Kenya (LSK) que tinha como objetivo a elaboração de um dicionário da língua de sinais. A autora relata sua experiência enquanto um exemplo de empoderamento da comunidade de surdos e o desenvolvimento educacional de uma pessoa surda local. A autora seguiu um dos princípios éticos fundamentais: compreender a língua de sinais de uma comunidade e identificar como estabelecer um benefício mútuo por meio de uma parceria com base no respeito entre o pesquisador e o pesquisado. Hochgesang partiu de Harris, Holmes e Mertens (2009) que apresentam questões éticas a serem consideradas no desenvolvimento de pesquisas com surdos e comunidades surdas.

Harris, Holmes e Mertens (2009) chamam a atenção para o fato de que comunidades surdas estão centradas em práticas linguísticas e culturais com as línguas de sinais, uma vez que a partir dessas línguas a coletividade surda se estabelece. As questões éticas discutidas pelos autores partem de três princípios básicos: respeito (tratamento das pessoas enquanto agentes autônomos), benefícios (segurança aos participantes quanto os benefícios para os próprios surdos e para suas comunidades, benefícios considerando seus interesses) e justiça (que a pesquisa seja algo que garanta a justiça na relação entre o pesquisador e o pesquisado e sua comunidade). O objetivo é engajar os participantes de um grupo culturalmente diverso considerando a justiça social. Quais os benefícios para os surdos e suas comunidades que garantem uma justiça social? Hochgesang (2015) formulou esta pergunta ao contribuir com a documentação da LSK. A autora assumiu uma posição de realizar a documentação colaborativamente com a comunidade surda envolvida, assim passou a contar com um colega surdo do Kenya enquanto pesquisador coautor na implementação da documentação desta língua. Assim, a proposta da pesquisa, a implementação e o uso dos dados coletados foram altamente sensíveis às demandas da própria comunidade. Logo, a relação da pesquisa com a comunidade e seus membros passou a desempenhar um papel empoderador (HARRIS; HOLMES; MERTENS, 2019), passando a ter um efeito transformador.

Hochgesang (2015) também mostrou o quanto faz-se necessário negociar entre o que os surdos locais querem e o que é possível fazer. Ela cita o exemplo do desejo dos surdos de terem um dicionário todo elaborado na LSK, mas que diante do contexto no qual as crianças surdas do Kenya estão inseridas, constatou-se a importância de incluir também a língua escrita do país, pois seus professores desconheciam a LSK. Isso foi discutido junto à comunidade que compreendeu essa necessidade, mas ficou, de certa forma, desapontada. No entanto, essas decisões foram feitas em conjunto com a comunidade. O pesquisador precisa perguntar à comunidade o que eles querem e pensar juntamente com eles sobre como viabilizar esses desejos, especialmente considerando todos os fatores contextuais que irão determinar como efetivamente será possível viabilizá-los. Após essa etapa, todo o processo de documentação precisa envolver surdos das comunidades em todas as etapas de execução. O projeto proposto de documentação da LSK sempre foi um projeto da comunidade surda do Kenya, assim o dicionário contou com surdos quenianos como coautores.

Há também um bloco de documentação de línguas de sinais para a constituição de *corpora* comparáveis entre variantes da mesma língua de sinais e entre diferentes línguas de sinais, uma vez que a metodologia da implementação destes *corpora* são similares. Os *corpora* constituídos estão disponíveis *on-line*:

Língua de Sinais Alemã (DGS) - <https://www.sign-lang.uni-hamburg.de/dgs-korpus/index.php>

Língua de Sinais Australiana (Auslan) – <http://www.auslan.org.au/about/corpus/>

Língua Brasileira de Sinais (Libras) – <https://corpuslibras.ufsc.br>

² Tradução nossa: “Quando pesquisadores ouvintes trabalham em estreita colaboração com pesquisadores surdos, a colaboração resultante pode trazer recompensas positivas; mas não vem sem preparação e esforço consciencioso”.

Língua de Sinais Britânica (BSL) – <https://bslcorpusproject.org/>
 Língua de Sinais Holandesa (NGT) - <https://www.ru.nl/corpusngten/>
 Língua de Sinais Japonesa (JSL) - <http://research.nii.ac.jp/jsl-corpus/public/en/index.html>
 Língua de Sinais Polonesa (PJM) - <https://www.plm.uw.edu.pl/>
 Língua de Sinais Sueca (STS) - <https://www.ling.su.se/teckenspraks-korpus>

Todos estes *corpora* se inspiraram no primeiro *corpus* de língua de sinais estabelecido, o *corpus* da Auslan. Atualmente, estes *corpora* já apresentam características diferentes, mas vários dados são comparáveis, como aspectos que compõem a metodologia e os tipos de textos disponíveis: entrevistas, conversas informais, narrativas e enunciação de lista de vocabulário, normalmente seguindo a proposta básica da Lista Swadesh. Outro ponto comum entre estes *corpora* envolve a disponibilização pública de, pelo menos, parte dos dados coletados para socialização nas comunidades surdas locais e internacionais. As questões metodológicas comuns incluem a coleta de dados feita por surdos locais das respectivas regiões do país, o que minimiza interferências de ordem sociolinguísticas. Estes *corpora* apresentam um objetivo mais acadêmico, mas mesmo assim vários deles procuraram congregar aspectos técnicos com aspectos socioculturais. Assim, os dados estão sendo anotados para sistematizar as buscas para fins de pesquisa linguística, mas também há socialização de vídeos para serem assistindo e usados pelas comunidades surdas e para outros fins, tais como, fins educacionais.

Entre estes *corpora*, exemplificamos este bloco com o *Corpus* da Língua de Sinais Sueca que se constituiu na Universidade de Estocolmo, no Departamento de Linguística. Esse *corpus* apresenta vários tipos de dados que são úteis para as pesquisas empíricas e para o ensino da STS, especialmente pelo banco de dados lexical e pelos *corpora*. Os objetivos do dicionário da STS é de oferecer uma obra o mais completa possível dos sinais desta língua. O grupo de lexicógrafos da seção de língua de sinais do Departamento de Linguística já documentou 19.000 sinais. O banco lexical está disponível online de forma pública desde 2008 e é atualizado desde então. A maioria dos sinais são acompanhados de produções contextualizadas, variações existentes dos sinais, uso de exemplos e fotos ilustrativas. Os principais critérios de busca incluem a palavra, a tradução, outros significados do sinal, sinais alternativos, configurações de mão e área de especialidade, quando é o caso (STS, 2020).

O primeiro *corpus* da STS foi estabelecido em 2004 no contexto do projeto ECHO (CRASBORN *et al.*, 2007). O banco de dados do *Corpus* da STS (2009-2011) consiste de 24 horas de vídeos (conversação, narrativas e apresentações pessoais) de 42 sinalizantes, entre 20 e 82 anos de idade de três regiões da Suécia. Os dados foram compartilhados como domínio público (ÖQVIST *et al.*, 2020) na página do *corpus* da STS (2020). Além destes *corpora* da STS, a Suécia conta com um *corpus* da Língua de Sinais Sueca Tátil, que consiste de diálogos e elicitación de produção com oito sinalizantes surdocegos (MESCH, 2016). Também, os *corpora* da STS contam com um *corpus* de STS como segunda língua com dados longitudinais coletados com 38 adultos ouvintes aprendizes de STS como L2 (SCHÖNSTRÖM; MESCH 2017, no prelo; LEESON *et al.*, 2019).

No caso do Brasil, o *Corpus* de Libras integra o Inventário Nacional de Libras que segue esta metodologia internacional para fins de pesquisa linguística, mas também socializa as produções que podem ser usadas em sala de aulas enquanto registro histórico das comunidades surdas brasileiras, assim como para fins literários da Libras. Além deste *corpus* mais formal, há também outros *corpora* da Libras que integram textos mais ou menos acadêmicos, assim como textos literários em Libras para serem utilizados com diferentes aplicações (ver aplicações QUADROS *et al.*, 2020).

É importante registrar que estes *corpora* precisam também servir para as comunidades surdas, uma vez que representam um registro histórico do estado sincrônico das respectivas línguas de sinais coletados em um momento específico que podem ser acessados pelos seus sinalizantes em diferentes momentos da história de sua língua, ou seja, podem servir para fins de pesquisa diacrônica no futuro. Também podem incluir produções em Libras que referendam as histórias surdas, portanto, servirem como produções de narrativas e de histórias nas línguas de sinais que podem ser usadas em sala de aula.

5 DIFERENTES TIPOS DE DOCUMENTAÇÃO LINGUÍSTICA DE LÍNGUAS DE SINAIS

As línguas de sinais são, ainda, um fenômeno humano pouco conhecido. Muitas pessoas pensam que elas são universais, ou que são uma codificação das línguas orais, ou que são inferiores a estas. No entanto, essas línguas circulam há muito em cada país onde surdos têm a possibilidade de viver em comunidade e de interagir entre si usando suas línguas de sinais. Desse modo, análogo ao que ocorre com as línguas orais, cada país tem sua própria língua sinalizada, a qual é manifesta por uma gramática com estrutura própria e marcada por fatores sociolinguísticos das comunidades de fala específicas desses países. Assim, não é porque na Inglaterra, Estados Unidos e Austrália se fala inglês que as línguas de sinais desses países são as mesmas (SCHEMBRI, 2010). Essa relação não é necessária. Sendo a língua um dos mais importantes e representativos elementos da cultura de um povo, atravessada por processos sócio-históricos, em cada país os surdos marcam sua identidade linguístico-cultural nas suas línguas de sinais, o que, em muitos casos, torna ininteligível a comunicação entre surdos de países diferentes.

Dada a diversidade linguística de línguas de sinais ao redor do mundo, bem como aos propósitos emergentes da Ciência, a documentação de uma língua pode ser feita de diversas maneiras, o que vai implicar diferentes tipos de documentação linguística. Cabe ao pesquisador delinear o tipo de *corpus* linguístico mais se adequado à pesquisa que ele quer empreender, sem, no entanto, esquecer aquele princípio básico para a realização de documentação linguística em comunidades surdas: a compensação (retorno) que a minha comunidade de informantes terá com a minha pesquisa e de que forma o surdo pode/deve protagonizar a pesquisa.

A documentação linguística de língua reflete, portanto, os objetivos que se quer alcançar. Assim, pode-se elaborar *corpora* dos seguintes tipos:

- a) *Corpus* Monolíngue - aqueles em que se usam apenas uma língua. Tipo de *corpora* que permitem ao usuário uma gama de ferramentas para manuseio dos dados, bem como verificar padrões e tendências da língua documentada.
- b) *Corpus* Paralelo (ou multilíngue) - aqueles constituídos por textos por dois (ou mais) *corpora* monolíngues, em que um texto é a tradução do outro. Nesse tipo de *corpora*, o usuário pode colocar um texto alinhado ao outro para fazer diferentes comparações quanto aos processos de segmentação e de análise linguística.
- c) *Corpus* Diacrônico - aqueles cujos registros datam de diferentes períodos de tempo e visam analisar as mudanças nas línguas.
- d) *Corpus* Comparável - aqueles formados por mais de dois *corpora* monolíngues, levando-se em consideração os mesmos metadados e princípios de elaboração, o que torna os resultados comparáveis.
- e) *Corpus* do aprendiz - aqueles formados a partir de produções linguísticas de aprendizes de uma língua estrangeira. Tem sempre por objetivo trabalhar aspectos relacionados ao ensino de línguas.
- f) Mistos - os que mesclam mais de um tipo de *corpora*.

Diante da diversidade de tipos de *corpora* que se pode elaborar, cabe a reflexão sobre o que se tem feito em línguas de sinais: a) quando se faz o uso de glosa para representar a produção linguística de um surdo captada em vídeo tem-se aí um *corpus* paralelo? b) Os *corpora* que envolvem tradução para a língua oral oficial do país de uma dada comunidade surda constituem *corpora* paralelos?

The screenshot displays the ELAN software interface. On the left, a video window shows a man and a woman in conversation. Below the video is a control bar with playback buttons and a timeline. The main area shows a detailed timeline with linguistic annotations in Portuguese and Libras. On the right, a 'UniSintática' panel lists syntactic units with their corresponding annotations.

N.	Anotação
60	[EL@ VIR EU]
61	DV
62	EU DV
63	EU OBRIGADO DV
64	DV
65	EU ASSUSTAR/
66	[EU CASA FAMILIA LINGUA DE SINAIS BOM, BRINCADEIRA
67	EL@ DV
68	[EU ESTRANHAR]
69	[EU ASSUSTAR]
70	[É PASSADO MOMENTO PERIODO COMUNICAÇÃO TOTAL ELA@]
71	[EL@ ORALIZAR]
72	[EU ASSUSTAR]
73	[EU FICAR-DE-PÉ ESTUDAR SALA
74	[EU VER SURDO [...]]*
75	[SINALIZAR BEM, NÃO.]
76	[TUDO BEM?]
77	[EU SURDO NAO]

Figura 2: Tela do ELAN com Sinal em Libras + Glosa em Português + Tradução para o Português

Fonte: Corpus de Libras da UFSC

Parece razoável não considerar que se tem na tela acima a presença de um *corpus* paralelo, uma vez que nem a glosa referente à unidade sintática [EU ASSUSTAR], tampouco a tradução [Eu ficava assustado] são decorrentes de um planejamento do pesquisador para que se possam comparar línguas diferentes (Libras/Português), mas apenas um recurso para busca de termos (glosa) e para compreensão daqueles possíveis pesquisadores que não sabem Libras o suficiente para compreender a produção em vídeo. Por essa linha de raciocínio, portanto, o *corpus* de Libras parece ser mais adequadamente classificado como monolíngue. Por outro lado, há muitas similitudes com a elaboração de outros *corpora*, a exemplo do *corpus* da Auslan, e é um modelo replicado em diversos estados brasileiros, o que o torna, também, um *corpus* comparável.

No geral, os *corpora* de línguas de sinais são do tipo monolíngues e comparáveis e a eles são vinculados recursos tecnológicos de busca e que visam à melhor compreensão dos vídeos, o que é feito por meio de glosas e da tradução, respectivamente. Pode-se dizer que a prática de elaboração de *corpora* em línguas de sinais é algo relativamente novo, mas um campo em expansão, o que significa dizer que se tem uma área de estudos em se constituindo e, portanto, com diversas possibilidades e desafios a serem enfrentados. Assim, *corpora* diacrônicos, paralelos, comparáveis (sobretudo entre línguas de sinais) e do tipo *corpus* do aprendiz devem ser feitos a fim de se empreender pesquisas que contemplem estudos históricos, comparados e voltados para o ensino de línguas.

6 CONSIDERAÇÕES FINAIS

Os desafios de se documentar uma língua de modo representativo por meio de um *corpus* teórico-metodologicamente consistente, levando-se em conta preceitos éticos, especialmente com comprometimento com a comunidade de informantes não é exatamente uma tarefa ordinária. Em se tratando de uma língua de modalidade visual-espacial, cujos informantes-alvo devem ser os surdos, a tarefa torna-se mais complexa, por um lado, e, por outro, extremamente necessária.

A documentação de línguas é necessariamente uma área multidisciplinar, em que diferentes áreas do conhecimento interagem: computação, linguística, história, cultura, antropologia, dentre outras. A elaboração de *corpora* em línguas de sinais atende, portanto, a propósitos que vão além da descrição gramatical dessas línguas. Por se tratar de línguas minoritárias, em conformidade com o que preconiza a UNESCO, são línguas em risco de extinção. Documentar é, nesse caso, uma estratégia de planejamento de políticas linguísticas capazes de fortalecer e disseminar as línguas de sinais, além de viabilizar seus status sociolinguístico e histórico.

Corpora que envolvam comparação entre diferentes línguas de sinais, produções literárias em diversos gêneros textuais, produções voltadas para ensino de línguas de sinais, produções de crianças surdas e codas são necessários para a ampliação e formação de pesquisadores na área de documentação de línguas de sinais. Tudo isso, necessariamente, incluindo o surdo nesses processos formativos e de registro da sua língua, tornando-o autor e ator dessa obra.

AGRADECIMENTOS

Este trabalho foi possível pelos recursos do Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPQ (440337/2017-8).

REFERÊNCIAS

- CRASBORN, O. *et al.* Sharing sign language data online: experiences from the ECHO project. *International Journal of Corpus Linguistics*, v.12, n.4, p.535-562, 2007.
- ELAN. *Sistema Eudico Annotator*. 2020. Disponível em: <https://archive.mpi.nl/tla/elan>. Acesso em: 16 abr. 2020.
- HARRIS, R.; HOLMES, H. M.; MERTENS, D. M. Research ethics in sign language communities. *Sign Language Studies*, v. 9, n.2, p.104-131, 2009.
- HOCHGESANG, J.A. Ethics of researching signed languages: the case of Kenyan Sign Language (KSL). In A.C. Cooper & K.K. Rashid (ed.). *Signed Languages in Sub-Saharan Africa: Politics, citizenship and shared experiences of difference*. Washington, DC: Gallaudet University Press, 2015. p.11-30.
- JOHNSTON, T. The reluctant oracle: annotating a sign language *corpus* for answers to questions we can't ask any other way. *Abstract of a Paper Presented at the Sign Language Corpora: Linguistic Issues Workshop*, London, 2009.
- JOHNSTON, T. The reluctant oracle: adding value to, and extracting value from, a signed language *corpus* through strategic annotations. *Corpora*, v.9, n.2, p.155-189, 2014.
- LEESON, L. *et al.* The uses of *corpora* in L1 and L2/Ln sign language pedagogy. In: ROSEN, S. R. (ed.) *The Routledge Handbook of Sign Language Pedagogy*. Routledge, 2019. p. 339-352.
- MCCARTHY, M.; O'KEEFFE, A. Historical perspective: what are *corpora* and how have they evolved? In: O'KEEFFE, A.; MCCARTHY, M.. *The Routledge Handbook of Linguistics*. New York: Routledge, 2010. p. 3-14.
- MESCH, J. Korpus för taktilt svenskt teckenspråk: Datamängd. [Tactile Swedish Sign Language Corpus]. *Section for Sign Language*, Department of Linguistic, Stockholm University, 2016.
- ÖQVIST, Z.; RIEMER KANKKONEN, N.; MESCH, J. STS-korpus : A sign language web *corpus* tool for teaching and public use. In: EFTHIMIOU, E. (ed.). Proceedings of the [LREC2020] 9th *Workshop on the Representation and Processing of Sign Languages: Sign Language Resources in the Service of the Language Community, Technological Challenges and Application Perspectives*. European Language Resources Association (ELRA), Marseille, France, 2020. p. 177-180.

QUADROS, R. M. de *et al.* *Corpus de Libras*. 2020. Disponível em: <http://corpuslibras.ufsc.br/>. Acesso em: 17 fev. 2020.

RAPPEN, R. Building a Corpus: What are the key considerations? *In: OKEEFFE, A.; McCARTHY, M. The Routledge Handbook of Corpus Linguistics*. New York: Routledge, 2010. p.31-37.

SCHEMBRI, A. Documenting sign languages. *In: AUSTIN, P.K (ed.). Language documentation and description*, v. 7. London: SOAS, 2010. p. 105-143.

SCHÖNSTRÖM, K.; MESCH, J. *Datamängd*. Projektet Från tal till tecken – att lära sig svenskt teckenspråk som andraspråk. [Dataset. The project From speech to sign – learning Swedish Sign Language as a second language], Department of Linguistics, Stockholm University, 2017.

SCHÖNSTRÖM, K.; MESCH, J. [no prelo]. Cross linguistic influence across modalities – the case of mouth actions in L2 sign language learners. *Sign Language Linguistics*.

SINGLETON, J. L.; JONES, G.; HANUMANTHA, S. S. Toward ethical research practice with Deaf participants. *Journal of Empirical Research on Human Research Ethics*, n.9, p.59-66, 2014.

SINGLETON, J.; MARTIN, A.; MORGAN, G. Ethics, deaf-friendly research, and good practice when studying sign languages. *In: ORFANIDOU, E.; WOLL, B.; GARY, M. (org.). Research methods in sign language studies: a practical guide*. Malden/USA – Oxford/UK: John Wiley & Sons Inc., 2015. p. 5-20.

STS. Disponível em: <https://teckensprakslexikon.su.se>. Acesso em: 17 fev. 2020.



Recebido em 24/09/2020. Aceito em 08/10/2020.