



OPEN ACCESS

Todo conteúdo de *Cadernos de Linguística* está sob Licença Creative Commons CC - BY 4.0.

EDITOR

– Ana Suelly Arruda Cabral (UnB)

AVALIADORES

– Jorge Lopes (UFPA)
– Marília Soares (Museu Nacional/UFRJ)

SOBRE OS AUTORES

- Ana Paula Brandão
Conceptualização; Recursos; Visualização; Escrita – Rascunho Original (§5); Escrita – Revisão e Edição.
- Patience Epps
Conceptualização; Recursos; Visualização; Escrita – Rascunho Original (§0, §6); Escrita – Análise e Edição.
- Susan Smythe Kung
Conceptualização; Recursos; Visualização; Escrita – Rascunho Original (§1); Escrita – Análise e Edição.
- Denny Moore
Conceptualização; Recursos; Visualização; Escrita – Rascunho Original (§2); Escrita – Revisão e Edição.
- Zachary O’Hagan
Conceptualização; Recursos; Visualização; Escrita – Rascunho Original (§4); Escrita – Revisão e Edição.
- Jorge Rosés Labrada
Conceptualização; Recursos; Visualização; Escrita – Rascunho Original (§3); Escrita – Revisão e Edição.

Recebido: 21/09/2022

Aceito 03/01/2023

Publicado: 02/05/2023

COMO CITAR

BRANDÃO, A.P.; EPPS, P.; KUNG, S.S.; MOORE, D.; O’HAGAN, Z.; ROSÉS LABRADA, J. (2023). Os acervos e a documentação linguística. *Cadernos de Linguística*, v. 4, n. 1, e666.



VERIFICAR
ATUALIZAÇÕES

RELATO DE EXPERIÊNCIA

OS ACERVOS E A DOCUMENTAÇÃO LINGUÍSTICA

Ana Paula BRANDÃO

Faculdade de Letras – Universidade Federal do Pará (UFPA)
Belém, Pará, Brasil

Patience EPPS

Department of Linguistics – The University of Texas at Austin (UT)
Austin, Texas, Estados Unidos

Susan Smythe KUNG

Archive of the Indigenous Languages of Latin America – The University of
Texas at Austin (UT)
Austin, Texas, Estados Unidos

Denny MOORE

Área de Linguística – Museu Paraense Emílio Goeldi (MPEG)
Belém, Pará, Brasil

Zachary O’HAGAN

California Language Archive, Department of Linguistics – University of
California, Berkeley (UC)
Berkeley, Califórnia, Estados Unidos

Jorge ROSÉS LABRADA

Department of Linguistics – University of Alberta (UAlberta)
Edmonton, Alberta, Canadá

RESUMO

Enquanto mais e mais línguas do mundo se tornam ameaçadas, sua documentação fornece recursos importantes para linguistas e comunidades. Os linguistas documentais consideram os acervos digitais como um recurso essencial para garantir a preservação, a conservação e o acesso aos resultados de seu trabalho. Neste artigo, consideramos os benefícios e desafios associados

ao arquivamento na documentação linguística, relacionados a questões de preservação, conservação, acesso, propriedade e uso de materiais. Temos por base nosso conhecimento acumulado como acadêmicos profundamente envolvidos na administração, uso e composição de acervos linguísticos, especialmente os relacionados às línguas indígenas da América Latina. Enfatizamos a relevância dos acervos linguísticos no Brasil e sua importância para acadêmicos, membros da comunidade e demais interessados. Nossa discussão considera os passos necessários para garantir a qualidade e longevidade dos recursos, os princípios e estratégias pelos quais os materiais arquivados podem ser disponibilizados, e as possíveis contribuições dos acervos a pesquisas em andamento sobre as línguas indígenas. Conforme apresentamos aqui, os acervos linguísticos fornecem recursos importantes para acadêmicos e comunidades que desejam revitalizar, manter ou simplesmente lembrar sua herança linguística e cultural.

ABSTRACT

As more and more of the world's languages become endangered, their documentation provides key resources for linguists and communities. Documentary linguists look to digital archives as an essential resource for ensuring the preservation, conservation, and access of the outcomes of their work. In this article, we consider the benefits and challenges associated with archiving in language documentation, relating to issues of preservation, conservation, access, ownership, and use of materials. We draw on our accumulated knowledge as scholars who are deeply involved in administering, contributing to, and using language archives, particularly relating to the indigenous languages of Latin America. We focus in particular on the relevance of language archiving in Brazil, and its significance for scholars, community members, and other stakeholders. Our discussion considers the steps that are needed to ensure the quality and longevity of resources; the principles and strategies by which archived materials may be made available; and ways in which language archives can inform ongoing work with indigenous languages. As we lay out here, language archives provide key resources for scholars and for communities who wish to revitalize, maintain, or simply remember their linguistic and cultural heritage.

RESUMO PARA NÃO ESPECIALISTAS

O registro e salvaguarda da riqueza linguística, cultural e histórica dos povos originários é uma importante estratégia para o fortalecimento e revitalização dessas línguas, que cada vez mais estão ameaçadas ou à beira do desaparecimento. Nesse artigo, apresentamos os avanços e problemáticas enfrentadas na administração e uso de acervos linguísticos, isto é, repositórios de materiais audiovisuais e textos de línguas indígenas com anotações linguísticas e outras informações, que são guardados e preservados por uma organização ou instituição para que possam ser reutilizados no presente e no futuro. Nosso objetivo é mostrar a relevância desses acervos linguísticos, principalmente no Brasil, além de discutir sobre princípios e estratégias para melhorar e facilitar o acesso a esses materiais para que os mesmos possam ser usados pela comunidade acadêmica e pelas comunidades indígenas.

PALAVRAS-CHAVE

Documentação Linguística; Arquivamento; Línguas Indígenas da América Latina.

KEYWORDS

Language Documentation; Archiving; Indigenous Languages of Latin America.

INTRODUÇÃO¹

À medida que mais e mais línguas do mundo se tornam ameaçadas, sua documentação fornece recursos importantes para acadêmicos e comunidades. Os materiais de documentação proporcionam uma base empírica para informar o nosso conhecimento sobre o que é possível na linguagem humana, um registro de diversas tradições culturais e discursivas e um registro tangível do patrimônio comunitário, oferecendo às gerações futuras acesso às vozes de seus pais e avós. No entanto, esses materiais tendem a ser frágeis e efêmeros. Fitas cassetes de áudio e vídeo quebram, cadernos mofam e desbotam, e até mesmo cartões SD e discos rígidos são suscetíveis a incêndios, inundações e mudanças tecnológicas, como evidenciado por eventos trágicos como o incêndio de 2018 no Museu Nacional, no qual inúmeras gravações e manuscritos valiosos foram perdidos. Cada vez mais, linguistas e outros profissionais consideram os acervos digitais como um recurso essencial para garantir a preservação e o acesso aos resultados do trabalho de documentação linguística.

Neste artigo, consideramos os benefícios e desafios associados ao arquivamento na documentação linguística. Nossa discussão se baseia em nosso conhecimento acumulado como estudiosos profundamente envolvidos na administração, uso e composição de acervos linguísticos, com ênfase nas línguas indígenas da América Latina. Nos concentramos particularmente na relevância do arquivamento de dados linguísticos no Brasil e sua importância para acadêmicos, membros de comunidades indígenas e outras partes interessadas.

Começamos esboçando os fundamentos da criação de acervos para iniciativas de documentação linguística: o porquê de o arquivamento ser tão importante, o que ele oferece, e como fazê-lo (§1). Com o objetivo de garantir a qualidade, longevidade e acessibilidade dos recursos, consideramos as melhores práticas contemporâneas em curadoria digital, as diferenças entre um acervo de línguas estabelecido e outras plataformas on-line, os benefícios do arquivamento e as decisões sobre o que e quando arquivar. Também exploramos a questão de onde arquivar, ou seja, determinar quais acervos estarão disponíveis e como eles serão organizados e mantidos, especialmente no caso das línguas indígenas brasileiras e de seus projetos de documentação. O acervo do Museu Paraense Emílio Goeldi fornece um estudo de caso instrutivo (§2). Nossa discussão então se volta para considerar a relação entre acervos e comunidades, com foco na ética do consentimento informado e questões de acesso comunitário à documentação, com base em um estudo de caso realizado por Jorge Emilio Rosés Labrada com falantes de Mako na Venezuela (§3). A importância do arquivamento de materiais de legado, resultantes de documentação realizada antes da era digital e frequentemente representada em mídias frágeis e com vida útil limitada, é abordada em §4. Por fim (§5), retorna-se ao contexto brasileiro com um estudo de caso detalhado de projetos de documentação e arquivamento realizados por Ana Paula Brandão com duas línguas

¹ Este artigo decorre de uma apresentação na Abralin ao Vivo, que aconteceu a 11 de julho de 2020. A apresentação gravada está disponível em <https://aovivo.abralin.org/lives/archiving-and-language-documentation/>.

indígenas do Brasil, Paresí-Haliti e Enawenê-Nawê, que destacam muitas das questões discutidas nas seções prévias. As observações finais são oferecidas em §6.

1. ARQUIVANDO DADOS DE DOCUMENTAÇÃO LINGÜÍSTICA

Esta seção apresenta as principais considerações relacionadas à importância e ao processo de arquivamento de dados de documentação linguística. Fornecemos uma análise geral da importância desta iniciativa (§1.1), seguida por uma visão geral dos tipos de repositórios digitais (§1.2) e dos benefícios do arquivamento para a documentação linguística (§1.3). Explicamos como os acervos digitais de línguas diferem de outras plataformas on-line (§1.4) e oferecemos algumas sugestões sobre como e quando os dados de documentação linguística devem ser arquivados (§1.5). Ao longo deste artigo, usamos o termo *dados* para nos referirmos a materiais de documentação linguística, isto é, registros e/ou observações de linguagem falada ou de sinais que podem ser processados, anotados e analisados (cf. GOOD, 2022). Os dados primários são as gravações brutas de áudio e vídeo ou registros escritos da língua, incluindo narrativas, histórias orais, elicitaciones, conversas, entrevistas e protocolos experimentais. Os dados secundários são transcrições, traduções, segmentações morfológicas, glosas e outros tipos de anotações que requerem algum nível de análise preliminar para serem criadas (HIMMELMANN, 2012; THIEBERGER E BEREZ, 2012; KUNG et al., 2020). Para acesso a uma visão geral do arquivamento linguístico e sua história, consulte também Henke e Berez-Kroeker (2016), Berez-Kroeker e Henke (2018), Kaplan e Lemov (2019) e Kung (2020).

1.1. A IMPORTÂNCIA DOS ACERVOS

Em seu artigo seminal que definiu o campo da documentação linguística, Himmelmann (1998) incluiu o arquivamento como um dos quatro passos-chave na criação do que ele chama de “uma documentação linguística” (p. 171, tradução nossa), ou seja, uma coleção de gravações audiovisuais transcritas e anotadas com os metadados que as acompanham. Woodbury (2003) chama essa coleção de dados primários e os metadados associados de *corpus*, e incluiu o arquivamento como um dos seis critérios que estabelecem a qualidade geral de um corpus documental.

Os *metadados* são a documentação contextual, técnica e administrativa de suporte que ajuda a explicar os dados, incluindo quaisquer chaves (por exemplo, códigos, ortografias) necessárias para compreendê-los, analisá-los e reutilizá-los (KUNG et al., 2020). A Tabela 1 (adaptada de CONATHAN, 2011) ilustra os níveis e tipos de metadados que são frequentemente solicitados pelos repositórios. As categorias na Tabela 1 não vêm de um acervo específico, nem refletem as categorias exigidas por todos os acervos. Em vez disso, elas constituem importantes metadados contextuais que os potenciais depositantes devem ter em mente ao organizar materiais para arquivamento.

(Para ler mais sobre este tópico, recomendamos NATHAN e AUSTIN, 2004; JOHNSON, 2004; CONATHAN, 2011; e KUNG et al., 2020.)

Metadados ao nível da coleção ou projeto	Metadados ao nível da sessão, gravação ou ítem	Informação biográfica sobre todos os colaboradores
<ul style="list-style-type: none"> • Nome, informações de contato etc. do pesquisador/doador principal • Descrição do projeto (escopo e duração) • Intervalo de datas de criação • Códigos ISO para as línguas • Nomes as línguas ou dialetos • Descrição da quantidade, formato, mídia • Descrição do equipamento usado na gravação • Resumo do conteúdo e sua importância • Identificação de materiais pessoalmente ou culturalmente sensíveis 	<ul style="list-style-type: none"> • Título e/ou breve descrição • Data de criação • Local de criação • Pesquisador ou autor principal • Cocriadores, colaboradores, consultores • ISO para a língua • Nome da língua ou dialetos • Descrição do conteúdo e importância • Formato ou mídia • Extensão física, tamanho do arquivo eletrônico, duração do áudio/vídeo 	<ul style="list-style-type: none"> • Nome completo • Sobrenomes (ou nomes/códigos anônimos), se relevante • Data de nascimento • Local de nascimento • Língua(s) principal(is) • Língua(s) secundária(s) • Função na criação do acervo

Tabela 1. Categorias principais dos metadados (adaptada de CONATHAN, 2011).

As principais razões para arquivar dados linguísticos são garantir sua longevidade e acessibilidade. Como discutiremos a seguir, os repositórios digitais oferecem opções de replicabilidade e proteção contra os riscos de incêndio, inundação, perda, mofo, insetos e demais ameaças à conservação de materiais físicos, como os vistos na Figura 1.



Figura 1. Cópias físicas de materiais de documentação linguística: caixas de papéis, pilhas de cassetes de áudio, cadernos manuscritos e fotocopiados (Pérez Báez: ailla:257492, DSC_0001; Pérez Báez: ailla:257543, DSC_0052; P. Epps. Todos os direitos das fotos estão reservados e os outros possuem permissão de uso).

Himmelmann (2006) expõe a importância de arquivar os dados primários que resultam de um projeto de pesquisa além de publicar os resultados analíticos do projeto, ou seja, a trilogia boasiana de gramática (esboço), dicionário e conjunto de textos (WOODBURY, 2003). Quando os dados primários são arquivados, eles podem ser reutilizados para trabalhos adicionais de documentação linguística, bem como em outros tipos de análise linguística. Além disso, arquivar os dados primários e secundários permite que os resultados analíticos possam ser verificados e reproduzidos (HIMMELMANN, 2006; BEREZ-KROEKER et al., 2018). Em última instância, arquivar dados

primários e secundários em um repositório digital onde eles sejam publicamente acessíveis facilita a reutilização desses dados e fornece um meio estável para citá-los na forma de um identificador persistente (para uma orientação recente, ver CONZETT e DE SMEDT, 2022). Identificadores persistentes, como Identificadores de Objetos Digitais (DOIs), *Handles* e *Uniform Resource Identifiers* (URIs) permitem que os criadores dos dados recebam o devido crédito por seu trabalho (BEREZ-KROEKER et al., 2018).

Os materiais de documentação linguística são coletados com grande esforço, tempo e (frequentemente) dinheiro; no entanto, eles podem representar algo muito mais profundo para os falantes: sua cultura, identidade e autodeterminação (AGNU, 2007; CARROLL et al., 2020). Assim, é extremamente importante garantir que os falantes tenham acesso à documentação. Idealmente, cópias da documentação devem ser deixadas na comunidade desde o início, mas, caso não seja possível, uma cópia deve ser repatriada ou devolvida à comunidade (KUNG, 2020; VAPNARSKY, 2020; R. MILLER, 2021; ver §3.3). Colocar os dados em um acervo digital é uma forma de repatriação digital (KUNG, 2020), supondo que as comunidades tenham acesso relativamente fácil aos dados. Embora essa não seja a única maneira de repatriar os dados para a comunidade de origem, é *uma* maneira de compartilhar esses materiais com a comunidade e, ao mesmo tempo, preservá-los para futuras gerações de falantes. No entanto, a realidade ainda é que muitos povos e comunidades indígenas ao redor do mundo não têm acesso (adequado) à internet. Nesses casos, um acervo local, regional ou nacional, uma biblioteca, um museu, uma escola, um escritório de governo ou algum outro local acessível à comunidade de fala deve ser identificado, e uma cópia dos dados deve ser depositada lá (WILBUR, 2014), além de serem depositados em um repositório digital mais amplamente acessível, que estará melhor equipado para preservar os dados digitais a longo prazo (ver abaixo). Independentemente de onde exatamente os arquivos são guardados, o compartilhamento e repatriação dos dados primários e secundários e das publicações resultantes deles apoiam os esforços de manutenção, recuperação e revitalização das línguas, disponibilizando materiais para falantes e seus descendentes. Embora esperemos que os dados possam ser usados para educação, manutenção ou esforços de revitalização dessas línguas, é impossível prever os usos reais que os falantes e seus descendentes farão dos dados (ver SPENCE, 2018; VAPNARSKY, 2020 e LUKANIEC, 2022 para exemplos recentes). No entanto, os dados contêm as línguas e herança cultural desses falantes, e eles merecem o direito de decidir como e quando usá-los (HOLTON et al., 2022).

1.2. REPOSITÓRIOS DIGITAIS

Ao longo deste artigo, usamos o termo *acervo* para nos referirmos a um repositório de materiais que são guardados e preservados por uma organização ou instituição, para que possam ser reutilizados no presente e no futuro. Um acervo pode conter materiais analógicos e/ou digitais (tanto criados digitalmente quanto digitalizados). Um *repositório digital* é um acervo de materiais ou registros

digitais, e um *acervo linguístico (digital)* é um acervo especializado em patrimônio cultural linguístico e materiais ou dados linguísticos. No campo da documentação linguística, a definição de *acervo* mais citada é a de Johnson (2004, p. 142; tradução nossa): “Um acervo é um repositório confiável criado e mantido por uma instituição com a demonstração de compromisso para com a permanência e a preservação a longo prazo dos recursos arquivados”. A definição de Johnson não menciona o componente digital, por isso oferecemos uma definição atualizada de Trevor Owens, chefe de gerenciamento de conteúdo digital da Biblioteca do Congresso dos EUA. Essa definição enfatiza o equívoco de achar que um software projetado para manejo de bens digitais é a mesma coisa que um repositório digital. Deste ponto em diante, usamos os termos *acervo* e *repositório* alternadamente. Owens (2018, p. 4; tradução nossa) escreve: “Um repositório não é um software [...] Um repositório é a soma de recursos financeiros, hardware, tempo da equipe e implementação contínua de políticas e planejamento para garantir acesso de longo prazo ao conteúdo. Qualquer sistema de software que se use [...] para preservar e fornecer acesso ao conteúdo digital é necessariamente temporário [...] provavelmente não durará para sempre [...] As instituições possibilitam a preservação” (ênfase no original). Existem alguns temas recorrentes nessas duas definições: instituição, compromisso, continuidade, longevidade e preservação. Abordaremos cada um desses temas nesta seção, mas a mensagem principal é que não se pode simplesmente construir um acervo digital e esquecer de sua manutenção. Um acervo é um *compromisso contínuo* por parte da instituição ou organização que decide criá-lo.

Felizmente, existem muitas instituições mundiais que assumiram o compromisso necessário para manter acervos digitais. A *Digital Endangered Languages and Musics Archives Network* (DELANMAN, <https://www.delaman.org/>) é uma rede de acervos associados que são especializados em dados de documentação linguística. Alguns desses repositórios são somente digitais, enquanto outros também possuem acervos físicos. Alguns se especializam em uma área específica do mundo (por exemplo, o Acervo de Línguas Indígenas de América Latina – AILLA, <https://ailla.utexas.org/> – cujo foco regional é indicado em seu nome), e outros aceitam material de qualquer lugar do mundo (por exemplo, o *California Language Archive*, CLA – <https://cla.berkeley.edu/>). Alguns acervos aceitarão apenas dados associados a um auxílio ou financiador específico (por exemplo, o *Endangered Languages Archive* – ELAR, <https://www.elararchive.org/> – aceita principalmente materiais coletados com o financiamento do *Endangered Languages Documentation Programme*, ELDP), enquanto outros aceitam principalmente materiais em ou sobre um idioma específico (por exemplo, o *Standing Rock Sioux Tribe Language and Culture Institute*, <https://dev-standing-rock.pantheonsite.io/collections>). Por conta do aumento da conscientização sobre o estado de ameaça da maioria das línguas do mundo, várias instituições latino-americanas, inclusive no Brasil, também se comprometeram a hospedar acervos de línguas locais; ver Seifart et al. (2008) para uma discussão aprofundada das fases iniciais deste processo. Infelizmente, muitos desses acervos não conseguiram continuar a longo prazo. Dois repositórios no Brasil, no entanto, continuaram operando e são discutidos em §3; são os acervos do Museu Paraense Emílio Goeldi ([DOI 10.25189/2675-4916.2023.V4.N1.ID666 Cad. Linguist., Campinas, V. 4, N. 1, 2023: 666](http://arqling.museu-</p></div><div data-bbox=)

goeldi.br) e do Museu do Índio (<https://www.gov.br/museudoindio/pt-br>). Outros acervos também foram criados mais recentemente, como o *Archivo de Lenguas, Culturas y Memorias Históricas del Ecuador*, que conta com o apoio institucional da FLACSO Ecuador (<http://languages.flacso.edu.ec/>)². Navegando pelos links acima, o leitor pode comparar as diferentes interfaces e possibilidades on-line oferecidas atualmente por esses repositórios.

Embora os acervos mencionados acima cubram uma ampla variedade de localizações geográficas e requisitos de coleta, eles podem não ser adequados para todos os projetos de documentação linguística. Acervos alternativos incluem outros tipos de repositórios de dados digitais, como os administrados por muitas universidades. Os conjuntos de dados linguísticos (mas não os dados primários da documentação linguística) podem ser enviados ao *Tromsø Repository of Language and Linguistics* (TROLLing, <https://site.uit.no/trolling/about/>). Dados primários e secundários, bem como conjuntos de dados, podem ser arquivados em repositórios de dados gerais, como Zenodo (<https://zenodo.org/>) e Harvard DataVerse (<https://dataverse.harvard.edu/>). Outras possibilidades podem ser encontradas no *Registry of Research Data Repositories* (<https://www.re3data.org/>).

1.3. VANTAGENS DOS ACERVOS LINGÜÍSTICOS DIGITAIS

Para pesquisadores envolvidos em projetos de documentação, há muitas vantagens em arquivar materiais de documentação linguística (dados). Isso inclui migração de formato simplificada, organização pessoal, acessibilidade, descoberta, acesso graduado, gerenciamento de direitos, colaboração, facilidade de citação e preservação digital de longo prazo.

Qualquer pessoa que já teve que copiar gravações de campo de uma mídia obsoleta (por exemplo, fitas cassete) para outro formato, que tentar abrir um arquivo antigo em uma nova versão do programa ou encontrar um programa que possa até abrir o arquivo antigo entende as dores de cabeça envolvidas na migração de formato³. No entanto, os pesquisadores que arquivam materiais de documentação linguística em um repositório digital nunca mais precisam se preocupar com a migração de formatos, pois ela faz parte dos fluxos de trabalho de arquivamento supervisionados pela equipe do acervo.

O arquivamento também facilita a organização pessoal. Logo depois de voltar do campo, a maioria dos pesquisadores ainda lembra onde estão localizados todos os arquivos que criaram e como estão organizados. No entanto, com o passar do tempo, a memória falha e fica muito mais difícil lembrar onde os arquivos estão, como estão organizados, como eles se relacionam e os metadados relevantes (ou seja, informações contextuais; ver Tabela 1) que permitirão sua reutilização futura, seja pelo pesquisador que os criou ou por outros. O problema é maior com os

2 FLACSO é a sigla de *Facultad Latinoamericana de Ciencias Sociales*.

3 Ver Han (2022) para uma discussão recente sobre as muitas maneiras através das quais os dados podem ser *transformados*.

arquivos digitais, que não podem ser facilmente distinguidos e, portanto, precisarão ser visualizados ou abertos para serem identificados. A solução é arquivar os dados da documentação da língua, bem como os metadados associados, o mais rapidamente possível após sua criação. Uma vez que os dados e metadados são arquivados em um repositório confiável, eles sempre podem ser encontrados e acessados novamente. Os pesquisadores não precisam mais lembrar onde os arquivos estão armazenados (por exemplo, em qual disco rígido externo, laptop ou sistema de armazenamento em nuvem) e não precisam mais fazer malabarismos para lidar com várias cópias e backups dos dados. O arquivamento ajuda a garantir que os dados e metadados não sejam perdidos, descartados ou esquecidos.

Uma vez que os materiais e metadados são arquivados, eles podem ser acessados por seus criadores, bem como pelos colaboradores, outros pesquisadores, membros da comunidade de fala e qualquer outra pessoa que possa precisar de acesso aos dados. Além disso, os acervos linguísticos digitais são construídos de forma a facilitar a *descoberta* de materiais, ou seja, a capacidade dos usuários de encontrar ou descobrir materiais que estão procurando. A capacidade de descoberta é relevante tanto para pesquisas dentro do próprio acervo (via pesquisa facetada ou direcionada, pesquisas avançadas, etc.), e para pesquisas realizadas por meio de plataformas on-line mais amplas, como a indexação do Google e a coleta de metadados pela *Open Language Archives Community* (OLAC).

Embora existam repositórios digitais que afirmam ser de acesso completamente aberto, todos os acervos linguísticos digitais têm regras sobre como os usuários podem interagir com os materiais, e a maioria dos acervos linguísticos digitais tem *acesso graduado*, pelo qual certos materiais podem ser restritos a usuários específicos e/ou condições de uso específicas (ver também §1.5). Em relação às regras de uso, todos os acervos linguísticos digitais têm algum tipo de termos ou condições com que os visitantes on-line devem concordar antes de acessar os arquivos de mídia. Muitos acervos digitais exigem que um usuário crie uma conta gratuita e faça login antes de poder acessar os arquivos. No entanto, as informações do catálogo (ou seja, os metadados) geralmente são acessíveis ao público, o que significa que qualquer pessoa que acessar a página da web do acervo pode acessá-las e lê-las. A maioria dos acervos linguísticos tem níveis de acesso aos arquivos de mídia, embora a maneira de que esses níveis de acesso funcionem varie muito de um acervo para o outro. AILLA usa níveis numerados que indicam graus de acesso; outros acervos têm funções de usuário especializadas e alguns usam códigos de cores para indicar quem pode acessar materiais específicos.

A maioria dos acervos digitais especializados em dados de documentação linguística lida com o gerenciamento de direitos de forma semelhante, embora os detalhes exatos variem entre eles. Na

maioria dos casos, os detentores dos *direitos originais*⁴ mantêm todos os seus direitos de propriedade intelectual e/ou cultural. Os titulares dos direitos concedem licenças não exclusivas ao acervo e aos utilizadores do acervo; os detalhes das licenças variam de um acervo para outro, e de acordo com as leis de direitos autorais específicas do país onde cada acervo está localizado. O uso comercial dos dados *nunca* é permitido.

O arquivamento de dados facilita a colaboração em vários níveis. Para pesquisadores que estão envolvidos em colaboração remota com a comunidade de fala ou com outros pesquisadores, ele pode ajudar toda a equipe do projeto a se manter organizada, à medida que o arquivamento é iniciado ou que as análises são concluídas. Além disso, uma vez que os dados são organizados e arquivados, eles podem ser descobertos por outros pesquisadores, o que, por sua vez, pode levar a novas oportunidades de colaboração. Ademais, o arquivamento de dados facilita a reutilização dos dados primários para diversos fins de pesquisa, incluindo uma grande quantidade de trabalho que está sendo feito nas áreas de processamento de linguagem natural e tipologia linguística. Finalmente, os dados arquivados podem ser citados, e isso é crucial para a criação de pesquisas reproduzíveis (BEREZ-KROEKER et al., 2018). Algumas revistas, como *Language Documentation & Conservation*, solicitam que os conjuntos de dados associados a um artigo sejam arquivados e citados adequadamente. Assim, os pesquisadores podem e devem citar seus próprios dados arquivados, e os pesquisadores que utilizam dados arquivados também devem citá-los.

Por fim, um repositório digital confiável tem um plano ativo para a *preservação digital de longo prazo* dos arquivos de mídia digital e metadados associados. De acordo com a *Digital Preservation Coalition* (DPC), a *preservação digital* "refere-se ao manejo de uma série de atividades necessárias para garantir o acesso contínuo aos materiais digitais pelo tempo que for necessário. [...] [Ela] refere-se a todas as ações necessárias para manter o acesso a materiais digitais além dos limites de falha de mídia ou mudança tecnológica e organizacional". A *preservação a longo prazo* é o "acesso contínuo aos materiais digitais, ou pelo menos às informações neles contidas, indefinidamente" (DPC, 2015; tradução nossa). Isso significa que a preservação digital é muito mais do que apenas fazer backup de arquivos. O trabalho de preservação digital inclui a migração de dados e metadados de um formato para outro e de um sistema de software para outro à medida que a tecnologia muda. Significa garantir redundância precisa de dados (o que significa que há cópias duplicadas armazenadas em vários tipos de mídia e em vários locais), e significa monitorar regularmente a integridade de todos os arquivos em todos os locais. De acordo com Owens (2018, p. 5; tradução nossa), "a preservação é o resultado do trabalho contínuo das pessoas e compromissos de recursos. O trabalho nunca termina [...] Não é algo que possa ser pensado como um custo único".

4 Os titulares dos direitos são as pessoas que têm o direito de reivindicar direitos autorais sobre uma criação ou reivindicar direitos de propriedade intelectual, propriedade cultural, direitos morais, direitos de execução ou outros direitos inerentes a uma obra ou criação em virtude de tê-la criado, associação/afiliação de grupo ou herança.

1.4. DIFERENÇAS ENTRE ACERVOS LINGUÍSTICOS E OUTRAS PLATAFORMAS ON-LINE

Agora que já discutimos as vantagens de arquivar dados de documentação linguística, queremos comparar os acervos digitais com outras plataformas on-line que são comumente confundidas com acervos digitais, como plataformas de compartilhamento de mídia social, sites da Internet e armazenamento de arquivos em nuvem.

Plataformas de compartilhamento de mídia social, como YouTube, Vimeo e SoundCloud (ver Figura 2), facilitam o compartilhamento de arquivos de vídeo e áudio com comunidades, falantes e outros, porque seu conteúdo é facilmente descoberto por meio de mecanismos de pesquisa on-line. Eles facilitam o gerenciamento de direitos, porque permitem que a pessoa que carrega o arquivo escolha entre os direitos autorais tradicionais ou a aplicação de uma licença *Creative Commons*⁵. No entanto, a maioria das pessoas não tem conhecimento suficiente dos direitos autorais tradicionais ou da aplicação das licenças *Creative Commons* para tomar uma decisão informada. Essas plataformas também têm suas próprias versões de níveis de acesso, podendo o conteúdo ser mantido privado ou tornado público.

Os *sites da internet* também facilitam o compartilhamento de conteúdo e um código pode ser adicionado para tornar as páginas detectáveis pelos mecanismos de pesquisa. Mesmo que seja possível limitar o acesso ao conteúdo, esse pode ser um processo tecnicamente desafiador. Os direitos autorais tradicionais se aplicam automaticamente aos sites, mas o desenvolvedor do site pode optar por aplicar Declarações de Direitos⁶ ou licenças abertas, como aquelas gerenciadas pela *Creative Commons*, às páginas da web e/ou aos arquivos vinculados.

Sistemas de armazenamento de arquivos em nuvem, como Dropbox, Google Drive, OneDrive e Box, são bons para compartilhar arquivos com seus colaboradores enquanto ainda se está trabalhando neles, bem como para controlar o acesso aos arquivos, mas eles não possuem a capacidade integrada de gerenciamento de direitos e não são detectáveis em pesquisas na web.

Embora todas essas opções sejam ótimas para compartilhar dados, e possam oferecer alguma forma de backup de conteúdo, nenhum desses sistemas está comprometido com a preservação digital de arquivos ou dados a longo prazo. Os termos de uso dessas plataformas dão o direito de descontinuar o serviço e excluir contas e seus conteúdos. Além disso, algumas plataformas, como o YouTube, retêm cópias de materiais mesmo quando os arquivos são excluídos por seus proprietários / publicadores, ou podem estar sujeitos a downloads automatizados por outros sites – um problema óbvio quando se trata de dados confidenciais (RICE, 2021).

5 As licenças *Creative Commons* são um tipo de licença aberta que funciona em conjunto com os direitos autorais tradicionais; ver <https://creativecommons.org/>, ou no Brasil: <https://br.creativecommons.net/>, para maiores informações.

6 <https://rightsstatements.org/en/>



Figura 2. Vídeo no YouTube de uma história contada em Caquinte por Antonina Salazar Torres (gravado por Zachary O'Hagan, arquivado permanentemente no California Language Archive, DOI: <http://dx.doi.org/doi:10.7297/X2Z60M7W>)

1.5. COMO E QUANDO ARQUIVAR DADOS DE DOCUMENTAÇÃO LINGUÍSTICA

Antes do século XXI, os resultados da documentação linguística eram tipicamente arquivados no final da carreira de um pesquisador ou mesmo após sua morte. No entanto, com o advento de equipamentos de gravação digital, de acervos linguísticos e de requisitos de arquivamento de dados, a maioria dos pesquisadores hoje não espera tanto tempo.

Hoje, o mantra dos arquivistas é "Arquive cedo e arquive com frequência!" Muitos acervos linguísticos recomendam *arquivamento progressivo* ou *incremental*. Sob esse modelo, um pesquisador ou equipe de pesquisa envia dados primários (por exemplo, gravações de áudio e vídeo e fotografias), sempre acompanhados dos metadados relevantes (como nomes dos participantes, data, local, línguas faladas e informações descritivas para contextualizar porque os arquivos de mídia foram criados; ver Tabela 1), para um repositório digital o mais rápido possível após a criação dos arquivos de mídia (ver etapa 1 na Figura 3, abaixo). Dados secundários (por exemplo, anotações, transcrições, textos glosados e interlinearizados) e análises ou resultados acadêmicos (etapas 2 e 3, respectivamente, na Figura 3) são adicionados posteriormente após a finalização (ROBINSON, 2006; NATHAN, 2013; KUNG et al., 2020). Na Figura 3, as etapas representam ondas de arquivamento para uma sessão de gravação. A etapa 1, que envolve o arquivamento dos dados primários e metadados relevantes, deve ser realizada assim que o trabalho de campo ou a fase de coleta de dados estiver concluída, ou mesmo enquanto essas atividades ainda estiverem em andamento. A etapa 2, arquivamento dos dados secundários e metadados relevantes, pode ser feita repetidamente, à medida que as transcrições e traduções vão sendo finalizadas. A etapa 3, arquivamento de análises finalizadas e produção acadêmica, juntamente com metadados relevantes, também pode ser feita de forma repetida e mais adiante no futuro. Essa abordagem reconhece que, para a maioria dos pesquisadores, ou qualquer outra pessoa, nunca haverá um momento realmente conveniente para arquivar dados de documentação linguística. Quanto mais tempo se passar e quanto mais arquivos digitais e materiais físicos se

acumularem entre a coleta de dados e o arquivamento, mais difícil e menos conveniente será o processo de arquivamento e mais tempo se levará para realizá-lo. Além disso, como tudo ainda está fresco na mente de um pesquisador imediatamente após um período de trabalho de campo ou coleta de dados, é muito mais fácil organizar nesse momento os materiais com precisão e garantir que os metadados estejam completos⁷.

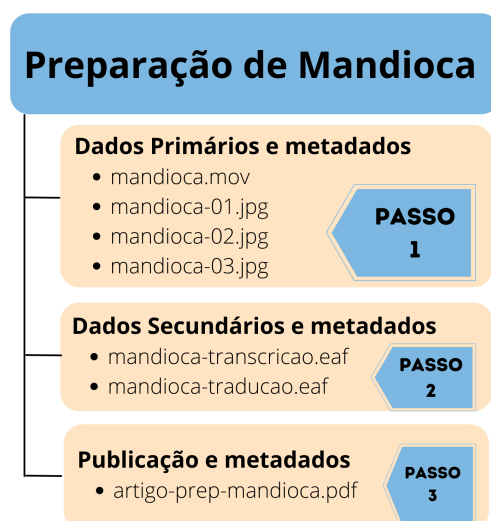


Figura 3. Arquivamento progressivo (gráfico de S. Kung).

Embora os acervos sejam fundamentais para a conservação, preservação, acessibilidade e descoberta a longo prazo dos dados de documentação linguística, nem todos os dados são igualmente adequados para arquivamento ou devem ser arquivados do mesmo modo. Alguns materiais podem ser culturalmente ou pessoalmente sensíveis ao ponto de simplesmente não poderem ser arquivados. Outros materiais podem exigir acesso restrito, conforme observado em §1.3 acima, de modo que apenas algumas pessoas possam ter acesso a eles. Às vezes, essas restrições podem se aplicar de maneiras diferentes a, por exemplo, um arquivo de áudio, um arquivo de vídeo ou uma transcrição escrita do mesmo evento. Exemplos de material potencialmente sensível incluem conhecimento esotérico ou protegido que não se destina a ser compartilhado com pessoas ou grupos de pessoas em particular (por exemplo, pessoas de fora da comunidade, membros de outros clãs, homens versus mulheres), discursos pessoalmente prejudiciais e informações que possam colocar em risco indivíduos ou comunidades. Documentação e arquivamento de línguas exigem comunicação aberta e contínua entre documentadores, falantes/sinalizadores de línguas orais/de sinais e outros

⁷ Para obter mais informações sobre como realizar o processo de arquivamento, também o processo de criar metadados, consulte 'Archiving for the Future: Simple Steps for Archiving Language Documentation' at <https://archivingforthefuture.teachable.com/>. As versões em português e espanhol deste tutorial estarão disponíveis no final de 2023.

membros da comunidade para garantir um processo totalmente ético e informado, conforme detalhamos na seção §3 abaixo.

Finalmente, conforme discutido nesta seção, ao se arquivar materiais de documentação linguística, é essencial escolher um repositório que garanta sua preservação e acessibilidade a longo prazo. Muitos acervos com a infraestrutura necessária, incluindo muitos do consórcio DELAMAN, atendem a uma comunidade internacional, e não devem ser vistos como proprietários em relação a qualquer país que possa hospedá-los. No entanto, as iniciativas de arquivamento são valiosas em todos os níveis: local, regional, nacional e internacional. Em alguns casos, pode ser aconselhável arquivar materiais em mais de um local, a fim de atender às prioridades locais e, ao mesmo tempo, aproveitar a infraestrutura de nível superior que oferece preservação e acesso mais confiáveis (ver §2.2 e §5, abaixo).

2. ACERVOS LINGUÍSTICOS NO BRASIL: O MUSEU GOELDI E OUTRAS INICIATIVAS

Nesta seção, examinamos mais de perto as iniciativas de arquivamento no Brasil e sua relevância para a documentação das línguas indígenas neste país. Dado o grande número de línguas indígenas no Brasil e a situação preocupante sobre o grau de vitalidade das mesmas (MOORE, GALUCIO e GABAS, 2008), a documentação linguística é urgente, além do fato de receber grande apoio entre os grupos indígenas. No levantamento das línguas do estado de Rondônia pelo Inventário Nacional de Diversidade Linguística, a documentação foi a segunda prioridade dos indígenas em relação às línguas faladas por esses grupos, com a correção de ortografias defeituosas e a produção de material escrito corretamente sendo a primeira prioridade (GALUCIO, MOORE e VAN DER VOORT, 2018, p. 217). Muitas comunidades indígenas têm jovens capacitados em informática, fato que facilita a documentação digital. Existe potencial para a produção de uma grande quantidade de gravações.

No que segue, descrevemos o acervo de registros de línguas indígenas do Museu Goeldi para apresentar em detalhes concretos e quantitativos um exemplo de um acervo situado no Brasil: como foi iniciado, criado, e o que contém (§2.1). Depois (§2.2), será oferecida uma breve apresentação da situação de outros acervos digitais de línguas no Brasil, entre os quais a do Museu do Índio⁸ é o exemplo principal. Também descrevemos alguns projetos de documentação de línguas e seus resultados, para indicar as necessidades de documentação no país, e discutimos a importância dos acervos regionais.

⁸ Esta apresentação constitui, portanto, uma atualização da discussão sobre os esforços incipientes de arquivamento no Brasil descritos em Seifart et al. (2008).

2.1. O ACERVO DO MUSEU GOELDI

O Museu Paraense Emílio Goeldi é um instituto de pesquisa do Ministério da Ciência, Tecnologia e Inovação (MCTI). Localizado em Belém, Pará, o Museu Goeldi foca em pesquisas sobre a Amazônia. A sua Área de Linguística, parte da Coordenação de Ciências Humanas, sempre foi precária. No ano de 1986, a Área não tinha nem gravador, nem computador e contava com apenas poucos livros. Como é frequente no país, havia pouco investimento em infraestrutura. Porém, existia uma respeitada tradição de coleções científicas no Goeldi e a documentação de línguas se encaixava bem nessa tradição. Estagiários, bolsistas e pesquisadores visitantes aumentaram as atividades da Área de Linguística, incluindo a documentação.

A infraestrutura começou a melhorar com a aquisição, através de projetos de pesquisa, de gravadores de fita cassete de boa qualidade, além de microfones, laptops e equipamentos de energia solar. Em 1996, o programa Centros de Excelência do Banco Mundial trouxe aparelhos de *Digital Audio Tape* (DAT), gravadores de vídeo Hi-8, uma ilha de edição Hi-8 e armários de armazenamento profissional. Com base em sua experiência acumulada, os linguistas do Goeldi, incluindo bolsistas e afiliados, concorreram com grande sucesso em programas internacionais de documentação interessados em apoiar projetos nos países em desenvolvimento. Nos primeiros anos da década de 2000, projetos do *Endangered Languages Documentation Programme* (ELDP) e do programa *Dokumentation Bedrohter Sprachen* (DoBeS) forneceram mais equipamentos e mais experiência. A documentação e aquisição de equipamentos (por exemplo, gravadores *solid state* com memória flash) continuaram na segunda metade da década, com apoio da USAID, UNESCO e o Fundo do Embaixador (Embaixada dos EUA). O MCTI forneceu equipamentos em 2007, incluindo um servidor como base para um acervo digital. Projetos de documentação (e pesquisa) continuaram na segunda metade da década de 2000, com apoio do CNPq, DoBeS e ELDP. Esse progresso foi ameaçado em 2007, por ataques fulminantes por parte de um pequeno grupo de linguistas, que fez oposição ao programa de documentação linguística do Museu do Índio, o ProDocLin (Projeto de Documentação de Línguas Indígenas, baseando no Museu do Índio⁹), aos programas internacionais de documentação e aos acervos digitais (MOORE e GALUCIO, 2016, p. 40). Felizmente, essa oposição não atraiu apoio entre os linguistas brasileiros e o progresso continuou, no Museu Goeldi e no Brasil.

Até 2009, as atividades dos linguistas do Goeldi resultaram em uma vasta coleção de diversas mídias de gravação: fitas cassetes, fitas DAT, minidiscos, fitas Hi-8, fitas mini-DV, etc. Por falta de tempo e orçamento, essa coleção foi simplesmente colocada em gavetas, sem nenhuma sistematização. No período de 2009 a 2014, a linguista Ana Vilacy Galucio coordenou um projeto, apoiado pelo Fundo de Defesa de Direitos Difusos do Ministério da Justiça, para digitalizar e catalogar a coleção, seguindo os padrões dos melhores acervos internacionais. Para agilizar o

9 <http://prodoclin.museudoindio.gov.br/>

processo de catalogação e armazenamento das gravações e seus metadados, *scripts* foram criados por Sebastian Drude e Rose Costa, reduzindo o tempo necessário em 70%. Assim, foi criado o acervo digital linguístico do Museu Emílio Goeldi.

Subsequentemente, mais equipamentos foram adquiridos e mais projetos de documentação conduzidos. Ao longo da evolução do acervo da Área de Linguística do Museu Goeldi, a tecnologia evoluiu e os formatos e mídias físicas também mudaram. Porém, a conversão de todas as gravações para um formato digital padronizado diminuiu os problemas com a longevidade e compatibilidade dos arquivos. Alguns aparelhos, como, por exemplo, microfones, não se tornam obsoletos. Já os vários tipos de gravadores atualmente sem uso (DAT, mini-DV, etc.) são guardados para ler gravações do passado e produzir um sinal que pode ser digitalizado. As gravações originais são guardadas nas suas mídias originais em armários profissionais.

Atualmente, as gravações em forma digital são armazenadas em um *Network Attached Storage* (NAS) de 32 TB. Para segurança, em caso de falha de um dos HDs do NAS, uma matriz redundante RAID-6 é implementada, reduzindo a capacidade de armazenagem a 22 TB. O conteúdo desse NAS é copiado no novo NAS de 96 TB, cuja capacidade é reduzida a 72 TB devido à redundância (ver Figura 4). Para garantir a segurança desses arquivos, em caso de incêndio ou raio, os dois NAS devem ser mantidos em prédios diferentes. As gravações de áudio são guardadas no seu formato original: .wav. As gravações de vídeo são guardadas no seu formato de maior definição e também em formato compactado (.mpg), que ocupa muito menos espaço. Esses arquivos compactados e os arquivos de áudio são catalogados e armazenados por meio do software *Language Archiving Technology* (LAT) (ver KOENIG et al., 2009). A maioria deles é copiada no servidor do Setor de Tecnologia da Informação e Comunicação do Museu Goeldi.



Figura 4. *Network Attached Storage* (NAS) de 96 TB no acervo do Museu Goeldi (foto: Denny Moore).

Atualmente, o acervo digital da Área de Linguística contém registros de 80 línguas indígenas. Desses, 73 são completamente digitalizados, catalogados e armazenados no software LAT. Os arquivos LAT ocupam 2,49 TB de espaço. A duração temporal dessas gravações é 1.561 horas de áudio e 474 horas de vídeo. Além desses arquivos, o acervo também

contém 9 TB de gravações brutas, incluindo vídeos de alta definição, que ocupam muito espaço. Trabalhos editados, principalmente para uso das comunidades, ocupam mais 1,5 TB. Outros 6 TB são ocupados por fotografias e arquivos avulsos dos pesquisadores. Os metadados dos arquivos LAT no servidor podem ser acessados no link <http://arqling.museu-goeldi.br>. As gravações não estão disponíveis para download atualmente, pendente da reorganização do acervo e da resolução de questões de acesso.

Para contribuir com o desenvolvimento da documentação de línguas indígenas no Brasil, os linguistas do Goeldi divulgam amplamente, através de treinamento intensivo, o conhecimento da tecnologia e metodologia envolvidas, além de oferecer sugestões dos melhores equipamentos. Esse treinamento ocorre inclusive entre as comunidades indígenas. O acervo também oferece serviços de digitalização e armazenamento para gravações de materiais de legado (ver §4, abaixo). Por exemplo, recentemente, houve um acordo com um antropólogo para digitalizar sua coleção de 115 fitas cassetes de dois dialetos Nambikwára. As gravações, algumas da década 1970, incluem música, um dicionário, conversas e pares mínimos de tom. As gravações estão sendo digitalizadas e, em colaboração com o antropólogo, logo serão catalogadas e depositadas no acervo linguístico digital. Ele receberá cópias no formato .mp3, que ocupam pouco espaço e podem ser devolvidas às comunidades Nambiwáras. As fitas cassetes podem ser depositadas no acervo ou retornadas ao pesquisador. Pessoas interessadas nesse serviço devem contatar a Área de Linguística do Museu Emílio Goeldi (linguistica@museu-goeldi.br).

2.2. OUTROS ESFORÇOS E PRIORIDADES DE ARQUIVAMENTO NO BRASIL

Além do acervo do Museu Goeldi, existem poucas outras iniciativas de arquivamento no Brasil. A principal delas está localizada no Museu do Índio, no Rio de Janeiro, que faz parte da Fundação Nacional do Índio. O Museu do Índio e o Museu Goeldi cooperam nos aspectos técnicos e metodológicos dos seus acervos. Ao longo do tempo, o Museu do Índio tem sistematicamente aumentado a sua infraestrutura, incluindo uma impressionante capacidade digital em armazenamento e backup. Além disso, o Museu do Índio vem conduzindo projetos ativos de documentação de línguas e culturas indígenas, seguindo os melhores padrões internacionais. O seu acervo digital contém aproximadamente 10 TB de documentação linguística, incluindo as línguas de 25 povos indígenas. Também contém 9 TB de documentação da cultura de 27 comunidades indígenas, o que inclui gravações de interesse linguístico. Materiais que resultaram do Projeto “Sonoridades Indígenas” de cinco povos ocupam 1,69 TB. O conteúdo total do acervo ocupa aproximadamente 50 TB (Thais Tavares MARTINS, Chefe do Serviço de Referências Documentais, Museu do Índio; c.p. a Denny Moore, 2020). Cópias de todas as gravações são devolvidas às comunidades envolvidas, mas o acesso via Internet ainda está sendo resolvido (ver <https://www.gov.br/museudoindio/pt-br>).

Alguns projetos de documentação linguística bem organizados estão produzindo quantidades significativas de gravações, aumentando a necessidade de arquivamento profissional. Por exemplo, o Centro de Documentação da Universidade Federal de Amapá, em Oiapoque, tem projetos de documentação com os Karipúna e Galibí-Marwóro (coordenado por Gelsama Mara Ferreira dos Santos) e com os Palikúr (coordenado por Elissandra Barros da Silva). Os dados estão atualmente sendo guardados em microcomputadores e HDs externos. O projeto com os Galibí-Marwóro tem 13,5 GB de áudio (.wav) e 403 GB de vídeo (.mov, .mp4), bem como 51 GB de fotos e 15 GB de vídeo editado. O projeto com os Palikúr tem 20 GB de áudio, 6 TB de vídeo (.mp4) e mais de cem mil fotografias no formato RAW, para um total de 10 TB (G. M. F. DOS SANTOS e E. B. DA SILVA, c.p. a Denny Moore, 2020).

Certamente, a demanda por acervos de documentação é crescente e muito acima da capacidade atual. Uma possibilidade para aumentar a capacidade é a criação de acervos digitais regionais. Esses acervos poderiam multiplicar a capacidade atual e teriam a vantagem da proximidade aos povos indígenas da região, facilitando a gravação e também o acesso dessas comunidades a esses arquivos. Um lugar natural para acervos deste tipo encontra-se nos programas de formação indígena nas universidades, por exemplo, nas Universidades Federais do Amazonas, de Goiás ou do Amapá em Oiapoque. Nestes programas, os alunos indígenas recebem formação em informática enquanto mantém contato com seus respectivos povos, o que facilita uma cooperação produtiva entre o acervo e as comunidades em questão. A infraestrutura desejável para um acervo deste tipo pode ser comprada de vez ou, como no caso do Goeldi, construída em etapas, em conjunto com projetos de pesquisa ou documentação. De fato, os componentes são relativamente acessíveis: O custo de um NAS (*Network Attached Storage*) de 48 TB era de aproximadamente R\$ 20.000 no meio de 2021 (com o dólar a R\$ 5,80). Para a documentação linguística digital, um kit completo de excelentes equipamentos semiprofissionais portáteis (filmadora, tripé, gravador de áudio digital, laptop, três tipos de microfones, lâmpada, maleta de transporte, baterias, cordas, adaptadores) custava aproximadamente R\$ 20.000. Seriam necessários pelo menos dois kits para equipar um acervo linguístico regional. Além da infraestrutura, é necessária formação em tecnologia e metodologia de documentação, para garantir qualidade, que é sempre um desafio. Uma questão a ser resolvida é a necessidade de softwares de arquivamento que sejam razoavelmente simples e de fácil uso.

Em última análise, as opções de arquivamento apresentadas nesta seção representam diferentes níveis de capacidade em relação aos objetivos principais de preservação, conservação e acessibilidade, de acordo com as observações apresentadas no final do tópico §1, acima. Embora as iniciativas de acervos em menor escala, como as propostas aqui, possam não atender aos padrões da DELAMAN com relação a alguns desses objetivos, elas certamente são um passo a frente e, em alguns casos, podem oferecer mais opções de acessibilidade, mesmo quando são menos desenvolvidas no que diz respeito a capacidade de preservação a longo prazo. Além disso, conforme observado acima, os pesquisadores devem estar cientes da possibilidade de arquivar tanto em nível

local ou regional quanto em um acervo de infraestrutura superior. Esta possibilidade é ilustrada pelos estudos de caso Paresí e Enawenê-Nawê, descritos em §5, abaixo.

3. ACERVOS E COMUNIDADES

Esta seção considera as maneiras pelas quais os membros de uma comunidade de fala podem se envolver em projetos de arquivamento e as questões éticas de consentimento informado e acesso que acompanham seu envolvimento.

Os membros das comunidades podem interagir mais ou menos diretamente com um acervo. Um exemplo de interação bastante direta pode ser visto no projeto "*Verdena Parker Collection of Hupa Sound Recordings and Films*" (PARKER, 2003+), mantido pelo *California Language Archive*. Verdena Parker, falante nativa de Hupa (Na-Dené, também conhecido como Atabascano-Eyak-Tlingit; Califórnia), fez gravações de som e filmes ao longo de muitas décadas, culminando em sua colaboração com linguistas da Universidade da Califórnia em Berkeley, em um projeto de documentação iniciado em 2005 (ver PARKER et al., 2005+). Essas gravações, arquivadas em 2010, incluem textos, vocabulário, traduções e observações sobre a vida no Hoopa Valley, para uso em programas de revitalização.

No entanto, a grande maioria dos projetos de arquivamento envolve pelo menos alguns falantes que não interagem diretamente com o acervo e, em muitos casos, podem não ter familiaridade prévia com o arquivamento ou com o que ele implica. O "Acervo de documentação da língua Kawahiva" (DOS SANTOS, 2017+) é um exemplo; esse acervo, desenvolvido de forma contínua, é composto por gravações em áudio e vídeo de histórias, conversas, canções, elicitções, encontros, notas de campo e fotografias. Dois falantes gravam, transcrevem e enviam arquivos mensalmente para arquivamento, "criando o sentimento de que o acervo também lhes pertence" (DOS SANTOS, c.p. a Zachary O'Hagan); e o depositante, Wesley dos Santos, também criou um vídeo para explicar aos membros da comunidade como acessar os materiais através do *California Language Archive*.

Qualquer que seja a afiliação da pessoa ou pessoas que lideram uma iniciativa de documentação e arquivamento, eles trabalham sob um imperativo ético para comunicar, explicar e obter o consentimento claro dos falantes e outros membros da comunidade envolvidos para realizar seus trabalhos. Abaixo, exploramos esses requisitos éticos e oferecemos algumas estratégias para atendê-los (§3.1), ilustrados por meio de um estudo de caso do trabalho do co-autor Rosés Labrada com o povo Mako da Venezuela (§3.2). No final desta seção (§3.3), consideramos estratégias para permitir o acesso da comunidade a materiais de acervo.

3.1. ÉTICA E CONSENTIMENTO INFORMADO

Discussões sobre ética e melhores práticas éticas têm figurado com destaque na literatura de documentação linguística. Nessas discussões, uma atenção especial foi dada à colaboração (GLENN, 2009; LEONARD e HAYNES, 2010), ao envolvimento e engajamento da comunidade (YAMADA, 2007; CZAYKOWSKA-HIGGINS, 2009; SAPIÉN, 2018; BISCHOFF e JANÝ, 2018) e à aplicabilidade de modelos éticos para diferentes partes do mundo (DOBRIN, 2008; HOLTON, 2009; PÉREZ BÁEZ, ROGERS e ROSÉS LABRADA, 2016). A relação entre arquivamento e ética também recebeu atenção considerável (por exemplo, Macri e Sarmento (2010) e Innes (2010), ambos exploraram questões éticas relacionadas aos acervos). Nessas discussões, o consentimento informado desempenhou um papel fundamental, com vários pesquisadores questionando o quão “informado” o consentimento realmente é (GRINEVALD, 2006; ROBINSON, 2010). Uma visão geral de algumas das questões relacionadas ao consentimento informado pode ser encontrada em Dwyer (2006, p. 43-48).

O consentimento informado é uma obrigação *legal* e *moral*, pelo menos em muitos lugares, incluindo o Brasil. É, portanto, essencial que os membros da comunidade compreendam o que a sua participação implica e, fundamentalmente, quais são realmente as consequências do arquivamento digital. Um desafio potencial ao consentimento informado em relação ao arquivamento é a familiaridade por parte da comunidade com os acervos e a Internet, e em alguns casos, computadores. Como indaga Robinson (2010, p. 189; tradução nossa), “podemos obter consentimento *verdadeiramente* informado [para arquivamento] se o consultor e a comunidade nunca viram um computador ou ouviram falar de um acervo ou da Internet?” Há duas questões adicionais aqui. Em primeiro lugar, a pesquisa é, por padrão, aberta. Geralmente, os linguistas reúnem um corpus baseado em questões de pesquisa específicas, mas essas questões, ou melhor, as respostas a elas, podem levar a novas questões. Assim, quando se obtém consentimento no início de um projeto, se os objetivos do projeto mudam ou surgem novas questões de pesquisa, o consentimento inicial ainda é válido? Isso é particularmente difícil de prever com dados arquivados que, como veremos abaixo, podem ser reutilizados por outros pesquisadores. Em segundo lugar, os descendentes e as comunidades do futuro podem precisar desses materiais para fins de revitalização, conforme descrito, por exemplo, em Bomfim (2017). A seguir, exemplificamos essas questões e propomos uma abordagem, entre muitas possíveis, a algumas dessas preocupações.

3.2. ESTUDO DE CASO: EXPLICANDO OS ACERVOS AOS MEMBROS DA COMUNIDADE

Este estudo de caso reflete sobre a experiência do co-autor Rosés Labrada ao explicar o arquivamento para membros das comunidades Mako ao longo do rio Ventuari na Venezuela durante seu projeto de doutorado (2012-2015) focado na documentação e descrição do Mako, uma língua Jodí-Sáliban (ROSÉS LABRADA, 2015).

Em 2011, Rosés Labrada realizou uma viagem a várias comunidades Mako ao longo do rio Ventuari e seus afluentes para obter o consentimento da comunidade para o projeto como um todo. Como parte desse processo, e atendendo aos protocolos locais, foram realizadas reuniões de aldeias em várias comunidades, que resultaram em um convite para voltar ao trabalho em duas comunidades: Arena Blanca e San José de Yureva. No entanto, nenhum dado foi coletado nessa primeira viagem. O projeto de documentação começou assim em julho de 2012, e Arena Blanca foi selecionada como o local da primeira estadia de campo devido à sua composição como uma comunidade falante de Mako amplamente homogênea¹⁰.

A preparação inicial para uma discussão com toda a comunidade sobre ética e consentimento informado ocorreu em 14 de julho de 2012 em uma reunião com dois membros da comunidade que também eram professores da escola em Arena Blanca na época e que concordaram em atuar como tradutores para a reunião. Um resultado claro da reunião com os professores foi a percepção de que, provavelmente, seria difícil para o linguista de fora explicar o conceito de *arquivamento* de modo que fosse claramente compreendido, devido à complexidade de alguns dos conceitos e tecnicismos por trás deles. Uma preocupação significativa foi o fato de que, na época, não havia conectividade para a Internet ou telefones, não havia computadores na comunidade e ambos os professores tinham pouca experiência com essa tecnologia.

A reunião da comunidade em 15 de julho contou com a participação de muitos adultos da aldeia, que se envolveram na discussão do projeto e nas possíveis questões éticas relacionadas à gravação e ao arquivamento de materiais audiovisuais. Por se tratar de uma reunião para pedir permissão para gravar, a reunião em si não foi gravada. No entanto, o processo de busca de permissão avançou significativamente durante este encontro, e permissões foram obtidas para:

1. realizar gravações em vídeo e áudio;
2. fotografar, mas com a ressalva de que fotos de crianças nuas ou mulheres sem blusa deveriam ser evitadas, pois a comunidade havia abandonado essas práticas tradicionais;
3. manter os dados além da duração do projeto, e compartilhá-los fora da comunidade.

No entanto, Rosés Labrada sentiu que o conceito de arquivamento permanecia “vago”, particularmente em duas áreas: (1) as implicações do compartilhamento on-line de materiais que identificam indivíduos específicos e (2) as opções de compartilhamento ou restrição de acesso a esses materiais ou aos nomes dos indivíduos. Assim, o linguista tomou medidas para garantir que a comunidade realmente entendesse o que estava envolvido no processo de arquivamento. Seu

10 San José de Yureva é uma comunidade mista Mako-Piaroa, onde tanto Mako quanto Piaroa são falados.

objetivo era mostrar como seria a coleção no acervo, mas o desafio veio da falta de acesso à Internet na comunidade.

A solução adotada foi fazer capturas de tela de um depósito inicial no acervo para mostrar como era o site e como seria navegar por ele. Em preparação, Rosés Labrada e um dos professores visitaram os sites da AILLA e ELAR enquanto estavam em Puerto Ayacucho, capital do estado do Amazonas, e o professor concordou em depositar uma de suas histórias gravadas para iniciar uma coleção que seria arquivada duplamente em ambos os acervos¹¹. Após o depósito de uma pasta com todos os itens correspondentes a uma única história, ambos os acervos forneceram capturas de tela de como era a coleção e o depósito.

Uma segunda reunião com a comunidade foi realizada em 4 de novembro de 2012 na Arena Blanca, após o retorno de Rosés Labrada com essas capturas de tela. A discussão que se seguiu foi melhor informada em vários aspectos. Primeiro, um pedido foi que os originais permanecessem na comunidade. Essa solicitação permitiu que o linguista fornecesse mais explicações sobre o contexto da documentação digital: que gravações específicas são criadas de maneira digital, que originais e cópias poderiam ser idênticos e que essas cópias poderiam ser feitas facilmente; o que não necessariamente teria sido o caso se Rosés Labrada tivesse gravado diretamente em fitas cassete ou CDs/DVDs, com os quais a comunidade estava mais familiarizada.

Uma segunda questão que surgiu foi em relação ao acesso para os membros da comunidade e a distinção entre os “que sabem”, ou seja, aqueles que sabem usar computadores e acessar a Internet, versus aqueles “que não sabem”. O consenso era de que aqueles que sabiam poderiam potencialmente acessar a Internet e usar esses materiais arquivados, mas para aqueles que não sabiam, seria necessário fazer cópias em DVDs e CDs. Uma terceira questão dizia respeito ao conteúdo das próprias gravações: uma membro da comunidade expressou uma preocupação em relação às gravações que poderiam conter piadas grosseiras ou palavrões e quem poderia ouvi-las. Para ela, não parecia ser um problema que pessoas de outras partes do mundo pudessem ouvir essas gravações. Em vez disso, ela estava preocupada em particular com as possíveis reações do povo Mako de outras aldeias próximas que poderiam ouvir e não entender que essas piadas eram realmente uma brincadeira. Por fim, foram discutidos os vídeos de atividades tradicionais como inofensivos, quando comparados a alguns outros vídeos que a comunidade conhecia, como filmes violentos. Isso proporcionou uma oportunidade para explicar ainda mais que falantes individuais tinham autonomia para restringir o acesso tanto no momento da gravação, quanto no futuro, e que as disposições específicas de acesso poderiam ser habilitadas e alteradas a qualquer

11 Tanto o AILLA quanto o ELAR concordaram inicialmente em hospedar os dados decorrentes do projeto de pesquisa. No entanto, ambos os acervos foram atualizados posteriormente para um modelo que evita o “arquivamento duplo” (ou seja, a duplicação das mesmas coleções em vários repositórios DELAMAN), pois isso significa o dobro do espaço de armazenamento digital necessário. A coleção Mako agora é hospedada pelo ELAR (<http://hdl.handle.net/2196/6bed9c49-c2dd-446d-b692-53c24cfbc916>) e o depósito inicial no site da AILLA (<https://ailla.utexas.org/es/isladora/object/ailla%3A124494>) redireciona o visitante para o site do ELAR.

momento no futuro. No geral, essa discussão garantiu a Rosés Labrada que os membros da comunidade da Arena Blanca tinham um entendimento mais claro (quando comparado à reunião inicial em julho de 2012) de algumas das questões éticas em torno do consentimento informado, arquivamento e acesso.

Este pequeno estudo de caso ilustra três pontos principais: 1) é crucial que nós, linguistas, expliquemos o arquivamento de maneira que as comunidades e os participantes individuais possam entender; 2) as comunidades e os participantes devem poder rever suas decisões; e 3) devemos tentar antecipar usos futuros dos materiais e fazer provisões sempre que possível, mesmo estando plenamente conscientes de que não podemos prever o futuro. Em última análise, as escolhas éticas não são necessariamente definidas de forma rígida, e o que funciona em um contexto específico pode não funcionar necessariamente em outro (HOLTON, 2009; GASSER, 2017). No entanto, desde que tenhamos como objetivo defender os princípios de respeito, reciprocidade, responsabilidade e relacionamento, conforme proposto por Rice (2006), devemos ser capazes de evitar algumas das possíveis armadilhas éticas que podem surgir como parte do processo de arquivamento.

3.3. PERMITINDO O ACESSO DA COMUNIDADE A MATERIAIS DE ACERVO

Como o caso Mako ilustra, ter e manter o acesso a materiais de documentação linguística tende a ser uma prioridade fundamental para as comunidades, não apenas para as pessoas que estão diretamente envolvidas na documentação, mas também, e às vezes até mais, para seus descendentes (por exemplo, DWYER, 2006, p. 59; VAPNARSKY, 2020; R. MILLER, 2021). No entanto, a devolução eficaz de materiais de documentação linguística às comunidades pode ser um processo complexo e de várias etapas, e deve levar em conta as diferentes capacidades que os membros da comunidade têm para acessar esses materiais. Essas capacidades podem ser limitadas não apenas pela falta de acesso à internet, por meio da qual os materiais arquivados em muitos repositórios digitais podem ser visualizados e baixados, mas também pelas habilidades necessárias para navegar nas interfaces on-line (bem como em relação aos idiomas em que essas interfaces podem ser apresentadas, como o inglês) e por conhecimentos básicos de informática. Para membros de algumas comunidades, o acesso até mesmo a tecnologias estáticas como os tocadores de CD e DVD pode ser limitado, assim como a capacidade de ler materiais impressos no idioma da comunidade (ou em outro idioma).

Habilitar o acesso pode exigir soluções criativas. Um modelo que obteve sucesso considerável é o “acervo audiovisual digital distribuído” ou “acervo jukebox” (BARWICK, 2004; BARWICK et al., 2005; O’MEARA e GONZÁLEZ GUADARRAMA, 2016). Essa iniciativa, originalmente testada na Austrália, envolve a instalação de um computador em um centro comunitário ou outro local neutro onde os membros da comunidade possam acessá-lo facilmente e não requer conexão com a Internet. O computador jukebox contém materiais de documentação linguística: gravações de áudio

e vídeo, bem como (potencialmente) transcrições, fotografias, etc. em formatos de arquivo facilmente móveis (como .mp3), além da capacidade de gravar CDs/DVDs e baixar arquivos em uma unidade flash, celular ou .mp3 player. Os membros da comunidade podem, assim, levar cópias de materiais para casa e também podem fazer o upload de seus próprios arquivos à jukebox. Outro modelo útil é um transmissor Wi-Fi local chamado "Raspberry Pi", que é eficaz em contextos em que as pessoas têm smartphones e as habilidades necessárias para esse uso, mas têm conhecimentos limitados ou inexistentes sobre computadores (por exemplo, THIEBERGER, 2019).

É claro que as considerações éticas são relevantes para todas essas iniciativas, e não apenas para o processo de habilitar e manter o acesso a materiais documentais, mas também para considerar como e se os membros da comunidade desejam limitar seu acesso e uso (ver as seções acima; também MACRI e SARMENTO 2010; DEBENPORT 2010). Alguns materiais gravados podem ser vistos como potencialmente prejudiciais, perigosos, ofensivos ou privados, e o entendimento de quem deve ou não ter acesso a quais materiais podem envolver categorias complexas e interseccionadas, associadas a concepções particulares de gênero, clã, parentesco, pertença à comunidade ou à vizinhança, etc., o que pode ser muito mais complexo do que o termo geral *comunidade* implica. Como apontado acima, as decisões sobre acesso aos materiais, como outras facetas da documentação e do arquivamento, devem ser continuamente informadas por princípios éticos.

4. MATERIAIS DE LEGADO E O ALCANCE DOS ACERVOS

Os acervos tendem a aumentar seu patrimônio com doações proativas de indivíduos vivos (geralmente pesquisadores acadêmicos) ou com legados de espólios de indivíduos falecidos. No entanto, o alcance dos acervos pode ser ampliado e enriquecido de várias maneiras para além desse modelo. Como é cada vez mais comum, falantes e outros colaboradores indígenas podem contribuir direta ou indiretamente para o processo de arquivamento, como expusemos em §3. Além disso, como expomos nas subseções abaixo, pesquisadores, membros da comunidade e funcionários do acervo podem ser proativos na localização e preservação de materiais antigos, um processo que pode envolver a colaboração entre muitas partes interessadas. Nesta seção, consideramos ideias práticas na localização e arquivamento de *materiais de legado*, isto é, materiais escritos ou audiovisuais coletados no passado, geralmente antes que os métodos de gravação e arquivamento digital estivessem disponíveis (§4.1) e em como fazer uso deles (§4.2).

4.1. VALORIZAÇÃO E ARQUIVAMENTO DE MATERIAIS DE LEGADO

Pesquisadores que trabalham ou trabalharam com falantes de línguas indígenas muitas vezes possuem um rico conjunto de materiais documentados, incluindo notas de campo, gravações de

som, fotografias e filmes, que podem estar armazenados em seus escritórios ou residências em condições precárias ou arriscadas (sujeitos, por exemplo, a danos por umidade, insetos, inundações). Esses materiais de legado, mesmo quando não são especialmente numerosos, são de alto valor linguístico, cultural, histórico e pessoal para indivíduos e grupos maiores. Para as línguas que não foram bem documentadas, materiais de legado podem constituir os únicos registros históricos do uso da língua, para não falar dos outros aspectos da vida que eles frequentemente registram. Ao contrário de línguas amplamente faladas, como o português, para as quais os registros históricos estão prontamente à mão, os registros históricos de muitas línguas indígenas geralmente persistem apenas por meio de esforços diligentes de arquivamento de materiais de legado. Como Austin (2017, p. 23, grifo nosso, tradução nossa) enfatiza: “Para projetos interessados em documentar, descrever ou revitalizar línguas, especialmente línguas ameaçadas, materiais históricos (digitais ou analógicos) como gravações em fita feitas em épocas anteriores, ou materiais escritos coletados anos ou mesmo séculos atrás, podem existir e podem representar importantes fontes de informação e, em alguns casos, a *única informação disponível*”. Os linguistas têm a responsabilidade profissional de estar cientes da existência de materiais de legado relevantes e, quando possível, trabalhar em sua preservação para e com as comunidades, para a Linguística e disciplinas relacionadas e para a posteridade em geral. Isso permite a posterior análise filológica do registro documental, “reconhecendo a filiação documental que caracteriza todos os dados linguísticos à medida que são sucessivamente registrados, interpretados e analisados” (GODDARD, 1973, p. 727; tradução nossa). Ao mesmo tempo, os linguistas têm a responsabilidade de garantir que seu envolvimento com materiais de legado seja realizado de forma ética (O’MEARA e GOOD, 2010; ver também §3 acima).

A preservação de materiais de legado envolve algumas etapas fundamentais: localizar os materiais; avaliar sua condição física; organizá-los de forma básica; transportá-los para um acervo ou outro local seguro (temporário); realizar qualquer remediação necessária para conservação; catalogá-los; e, idealmente, digitalizá-los (talvez não necessariamente nesta ordem). Localizar materiais de legado é muitas vezes uma das etapas mais difíceis, exigindo que o linguista seja informado ao acaso de sua existência e localização, ou que seja feita uma ampla investigação. Em geral, a localização de materiais de legado é facilitada ao se ter uma profunda familiaridade com as histórias de pesquisa e demais atividades de outras disciplinas, como explorações realizadas em regiões específicas. Mais concretamente, deve-se prestar atenção às descrições, muitas vezes na metodologia ou seção similar, das primeiras publicações (e trabalhos inéditos, quando disponíveis).

Por exemplo, o falecido antropólogo Gerald Weiss (para maiores informações, ver O’HAGAN, 2021), em uma seção de sua tese de doutorado intitulada “Design and Method”, disse o seguinte sobre o resultado de seu trabalho de campo nas comunidades Ashaninkas ao longo do rio Tambo no Peru entre os anos de 1961 e 1964 (WEISS, 1969, p. 6; tradução nossa):

As técnicas empregadas em campo para obtenção de informações eram padronizadas e ajustadas apenas às necessidades particulares das informações coletadas. Um diário foi mantido; leituras de temperatura, umidade e precipitação foram registradas diariamente no acampamento; as informações obtidas na observação e interrogatório foram acumuladas em fichas de quatro por seis ou outra forma conveniente,

como duplicatas com data, local e nome do informante indicados em cada ficha; espécimes, fotografias e gravações foram obtidas de tudo o que era possível.

O co-autor O'Hagan estava familiarizado com o trabalho de Weiss no contexto de seu próprio trabalho de campo com falantes de uma língua relacionada (Caquinte), mas, em leituras anteriores da tese, havia pulado a seção "Design and Method". Por meio desta seção, entretanto, ele ficou sabendo da existência (pelo menos em um determinado momento) de notas de campo, fichas, espécimes biológicos, fotografias e gravações de som. Gerald Weiss tornou-se professor de antropologia na Florida Atlantic University; uma pesquisa no Google em março de 2021 revelou que ele ainda era afiliado a essa instituição como professor emérito. Além disso, um telefonema ao mesmo tempo revelou que todos os materiais listados acima, e mais resultados da pesquisa de pós-doutorado, foram mantidos em sua casa na Flórida (para mais detalhes, ver ANWAR, 2021).

A organização básica de materiais de legado pode ser difícil se os objetos em questão (por exemplo, notas de campo, fitas) tiverem poucos metadados associados a eles e se especialistas que possam ler ou entender a versão falada de uma determinada língua não puderem ser consultados. Além disso, alguns arquivos podem estar em um formato que não é de fácil acesso, como é frequentemente o caso das fitas analógicas bobina a bobina, que eram comuns entre os anos 1950 e 1970, um período especialmente valioso na documentação de muitas línguas. Dependendo da situação específica, nesse momento, pode ser benéfico colaborar com membros da comunidade que possam ajudar na interpretação dos materiais, ou primeiro digitalizá-los, seja com auxílio profissional de um técnico de áudio externo¹² ou de um acervo como parte de uma doação. Independentemente disso, o objetivo principal desta etapa de organização é produzir um inventário básico de objetos (quantos cadernos, fitas, etc.) antes de serem transportados, para que haja um registro que possa ser verificado e garantir que todos os materiais cheguem no destino pretendido. Se faltar metadados, outras descrições dos objetos físicos podem ser fornecidas (por exemplo, "duas fitas cassete com desenho vermelho na capa").

O objetivo no transporte de materiais de legado é que os arquivos sejam enviados para um local que (em caráter temporário) tenha condições de armazenamento relativamente melhores (por exemplo, um local menos úmido) ou facilite a transferência para um acervo, caso não seja enviado para um acervo diretamente. O momento do transporte pode ser um dos mais perigosos para os materiais de acervo. Idealmente, materiais especialmente delicados como fitas podem ser transportados pessoalmente, por exemplo, na bagagem de mão¹³. Para outros materiais, é importante mantê-los secos (por exemplo, com coberturas plásticas) e separados de outros que possam danificá-los. Se os materiais forem enviados por correio, certifique-se de usar caixas de

12 Os estúdios de gravação mais antigos em áreas urbanas geralmente têm o equipamento e experiência em fazer essa digitalização. Existem vários programas de software para digitalizar o som analógico, como o Audacity ou o Sound Forge. Quando não for possível operá-los e houver financiamento suficiente, pode-se consultar um profissional.

13 Deve-se evitar, a todo custo, a exposição a ímãs, que podem danificar gravemente as fitas analógicas.

papelão de alta qualidade (com bastante fita adesiva e plástico bolha) e, quando financeiramente possível, opte pelo correio aéreo em vez do transporte terrestre, pois geralmente envolve menos manuseio das caixas. Em geral, nesta fase, o melhor é entrar em contato com um acervo que possa fornecer orientações adicionais. Os acervos também disporão de melhor conhecimento sobre catalogação e digitalização. Ressaltamos que a digitalização de gravações de som analógico e filme é, na maioria das situações, relativamente urgente, pois a fita original se degrada com o tempo e, na maioria dos casos, as condições de armazenamento prévio dos materiais não estão do mesmo nível das de um acervo. A umidade e os danos causados pela água são especialmente comuns.

4.2. UTILIZANDO MATERIAIS DE LEGADO

Os materiais de legado geralmente preservam as vozes e o conhecimento de pessoas que atuaram décadas antes de outros projetos de documentação e que, em alguns casos, representam os últimos falantes ou semi-falantes da língua como língua materna. Nesses casos, os materiais de legado podem ser uma fonte importante de informação sobre os esforços da comunidade na revitalização da língua e na manutenção ou recuperação do patrimônio cultural, bem como sobre os esforços dos acadêmicos em entender a amplitude e a profundidade da expressão humana.

Uma vez que os materiais tenham sido arquivados, muitos anos podem passar antes de serem utilizados por outros. Por exemplo, a linguísta Catherine Callaghan fez uma série de gravações de Sarah Ballard falando Bodega-Miwok (Miwokan; Califórnia) em 1960, enquanto a primeira era estudante de pós-graduação em linguística na Universidade da Califórnia em Berkeley. Ela doou essas e outras gravações para o *California Language Archive* em 1979 (BALLARD e CALLAGHAN, 1960)¹⁴. Quarenta anos depois, em 2019, o linguísta Andrew Cowell fez transcrições alinhadas ao tempo dessas gravações no ELAN, que ele arquivou no mesmo repositório (COWELL, 2019+)¹⁵. As duas coleções estão vinculadas ao catálogo digital do arquivo e, portanto, podem ser facilmente relacionadas entre si ao consultar uma ou outra. Da mesma forma, o linguísta Gladwyn Kingsley Noble Jr. fez gravações de falantes de Wapixána e Atorada (Aruak; Brasil, Guiana) durante uma única viagem de campo à Guiana em 1965. Depois de passar pelas mãos de diferentes acadêmicos, as 13 fitas foram doadas ao *California Language Archive* por volta de 2006 por Manjari Ohala (ver GEORGE et al., 1965). As gravações do Atorada, que antes se pensava não ter mais falantes nativos, foram utilizadas para uma descrição fonológica preliminar da língua por O'Hagan (2018), que foi elaborada por E. Miller (2021) como parte de um trabalho de conclusão de curso de graduação. Em 2021, os textos incluídos nas gravações foram traduzidos por falantes remanescentes que agora

14 Na época, o *California Language Archive* (CLA) era conhecido como *Survey of California Indian Languages*.

15 No modelo de arquivamento progressivo, em 2021, Cowell adicionou transcrições alinhadas no tempo às gravações de Richard Applegate com Sarah Ballard (BALLARD e APPLGATE, 1974).

residem em comunidades Wapixána, e serviram de base para um workshop dedicado à língua Atorada no final de 2022 (K. RYBKA, c.p. a Zachary O'Hagan).

Por fim, ressaltamos que os materiais de legado podem ser utilizados continuamente, e que o uso requer interpretação, isto é, um processo criativo de compreensão de "o que esses materiais significaram para seus criadores, que novos significados podem assumir no contexto em que estão sendo usados, e quais papéis [agentes contemporâneos] enquanto indivíduos podem desempenhar na circulação e recepção dos materiais" (DOBRIN e SCHWARTZ, 2021, p. 23; tradução nossa). Um exemplo especialmente produtivo de trabalho contínuo com materiais de legado é a bienal *Breath of Life Archival Institute for Indigenous California Languages* (ver GEHR, 2013) realizada na Universidade da Califórnia em Berkeley e organizada pela *Advocates for Indigenous California Language Survival* (AICLS) em conjunto com o *Survey of California and Other Indian Languages*, que abriga o *California Language Archive*. O Instituto traz pessoas indígenas ao campus de Berkeley para colaborar com linguistas voluntários na interpretação de materiais do acervo, muitos dos quais datam do início do século XX. Esse modelo já foi expandido com sucesso para outras localidades (BALDWIN, PÉREZ BÁEZ e HINTON, 2018) e pode ser proveitoso considerar se uma iniciativa semelhante poderia aumentar a acessibilidade dos materiais linguísticos existentes no Brasil.

5. ESTUDOS DE CASO: DOCUMENTAÇÃO E ACERVOS DAS LÍNGUAS PARESÍ-HALITI E ENAWENÊ-NAWÊ

Esta seção oferece dois estudos de caso que ilustram a documentação e o arquivamento linguístico no Brasil, por meio de projetos realizados em colaboração com os povos Paresí-Haliti e Enawenê-Nawê e liderados pela co-autora Ana Paula Brandão. Após uma breve introdução a essas duas línguas, descrevemos os projetos de documentação (§5.1), e as coleções de acervo resultantes (§5.2).

O Paresí (código *Glottolog* pare1272, latitude -14,59 e longitude -57,41) é uma língua indígena falada por um povo de mesmo nome, cuja população é de aproximadamente 3000 pessoas distribuídas em várias comunidades no Estado do Mato Grosso, perto da cidade de Cuiabá, nos afluentes do rio Juruena. Já a língua Enawenê-Nawê (código *Glottolog* enaw1238, latitude -12,43 e longitude -58,98) é falada por um povo que possui uma população menor, de aproximadamente 1000 pessoas, que vivem em duas comunidades (*Halataikwa* e *Kolinakwa*), em Terra Indígena localizada próxima às cidades de Juína, no Estado do Mato Grosso, e Vilhena, no Estado de Rondônia.

As duas línguas pertencem à família Aruak (PAYNE, 1991; AIKHENVALD, 1999; RAMIREZ, 2001). Com relação ao subagrupamento, Brandão, Carvalho e Pereira (2018) e Pereira (2018) apresentam evidências de que o Enawenê-Nawê é uma língua geneticamente muito próxima da língua Paresí,

além de proporem a existência de um subgrupo chamado *Juruena*, que incluiria as referidas línguas juntamente com a língua saraveca. Nas classificações mais importantes sobre a família Aruak, a única indicação de proximidade entre as duas línguas foi apontada em Fabre (2005); Payne (1991) agrupou o Paresí junto com a língua Waurá no grupo central e Aikhenvald (1999) também classificou o Paresí juntamente com as línguas do Xingu, formando um ramo chamado *Paresí-Xingu*.

Há uma documentação e descrição considerável da língua Paresí. Dois pesquisadores, a co-autora Ana Paula Brandão e Glauber Romling da Silva, de forma independente, realizaram a documentação da língua por vários anos. Entre os principais trabalhos de descrição linguística, estão Silva (2009, 2013) e Brandão (2010, 2014). Já o trabalho com os Enawenê-Nawê é bastante recente. Os únicos trabalhos descritivos conhecidos até o momento são os de Rezende (2003, 2013¹⁶), Brandão e Reis (2020), Reis (2020). Em 2019, Brandão recebeu financiamento do ELDP, por meio da Universidade Federal do Pará, para o projeto “Documentação da língua Enawenê-Nawê”.

5.1. OS PROJETOS DE DOCUMENTAÇÃO

O projeto de documentação da língua Paresí iniciou em 2006 e foi desenvolvido durante pesquisa de doutorado e pós-doutorado sobre a língua realizada por Brandão. Já o projeto de documentação da língua Enawenê-Nawê teve início em maio de 2019 e foi concluído no final de 2022. Ambos os projetos visam organizar um grande corpus de gravações em áudio e vídeo de um conjunto variado de gêneros linguísticos, que foram transcritos nas línguas indígenas e traduzidos para português.

A documentação Paresí foi feita inicialmente a pedido da comunidade Rio Formoso, que estava interessada no registro de sua cultura tradicional. Na comunidade Halataikwa dos Enawenê-Nawê, também entramos em contato com um falante da língua que nos convidou para visitá-los e apresentar a proposta do projeto para eles. Obtivemos autorização das comunidades para a gravação de material, sem fins lucrativos. As comunidades foram beneficiadas com a produção de DVDs, CDs e pendrives com histórias e canções tradicionais e com a formação de professores indígenas para a documentação linguística. O material Paresí serviu ainda como base para a elaboração da gramática de referência da língua, que foi defendida como tese de doutorado (BRANDÃO, 2014).

Entre os participantes do projeto, estavam professores indígenas e anciões, conhecedores da cultura tradicional. Alguns professores trabalharam na gravação, outros na transcrição e tradução e organização de metadados; um dos consultores Paresí também trabalhou com edição de vídeos. A presença do colaborador Paresí, Jurandir Zezokiware, na comunidade dos Enawenê-Nawê foi muito importante para estabelecer uma relação de confiança com os Enawenê-Nawê. Ele auxiliou durante os treinamentos e coletou dados linguísticos. Também foi possível contar com a participação de

¹⁶ Rezende (2013), uma descrição morfosintática da língua, não está disponível para a comunidade acadêmica nem para a comunidade de falantes.

alunos de graduação da UFPA, que visitaram as comunidades Paresí e Enawenê-Nawê a fim de aprender sobre o trabalho de campo.

Durante os projetos, foram adquiridos equipamentos de alta qualidade para os Paresí e Enawenê-Nawê através de apoio financeiro do ELDP. Utilizamos gravador digital Zoom, microfones de cabeça Shure, câmera de vídeo digital, microfone externo para filmadora, tripé, entre outros. No projeto com os Paresí, foram documentados vários eventos culturais e de fala, escolhidos pelas comunidades (por ex.: as histórias tradicionais, as músicas, as benzeções, as festas tradicionais, os jogos indígenas, entre outros). Além disso, registramos as diferentes variedades da língua Paresí. Os Enawenê-Nawê preferiram focalizar o trabalho na gravação de histórias tradicionais (por ex.: a história da origem dos Enawenê-Nawê, a história da origem da mandioca, histórias sobre os espíritos, entre outras). Mais informações sobre os dois projetos estão contidas em Brandão e Zezokiware (2018).



Figura 5. Os Paresí com as vestimentas tradicionais (PAB-200712-AP-RC-PontePedra80.JPG, foto: Rose Costa) e os Enawenê-Nawê contando histórias (UNK-20190500-AP-treinamentos-136.jpg, foto: Ana Paula Brandão).

5.2. OS ACERVOS DAS LÍNGUAS PARESÍ E ENAWENÊ-NAWÊ

Os dados primários, em áudio e vídeo, foram gravados nos formatos .wav e .mts, respectivamente. Cada sessão de gravação foi registrada em áudio e vídeo. Posteriormente, para a organização no acervo, os arquivos .mts foram convertidos para .mp4, por ser um formato mais compacto para armazenamento. Os arquivos estão nomeados com os códigos ISO 639-3 para as línguas (PAB para Paresí e UNK para Enawenê-Nawê), seguido da data de gravação (no formato AAAAMMDD), das abreviaturas dos nomes da pessoa que gravou e do entrevistado, além da palavra-chave da sessão na qual o arquivo será incluído, por exemplo: UNK-20200114-WE-YI-Kolito.mts. Como resultado dos projetos de documentação temos arquivos digitais com mais de 150 horas de gravações Paresí e 37 horas de gravações Enawenê-Nawê, que estão organizadas em sete categorias, como ilustrado na Tabela 2.

Categorias	Paresí	Enawenê-Nawê
Atividades do cotidiano	9h	30 min
Atividades rituais	13h	0
Elicitação geral	36h	30 min
Elicitação lexical	10h	4h
Músicas	15h	1h
Histórias tradicionais	38h	24h
Histórias não-tradicionais	30h	7h
Total	150h	37h

Tabela 2. Arquivos digitais de Paresí e Enawenê-Nawê.

Já os dados secundários incluem as anotações das gravações e os metadados. As anotações (transcrições e traduções para o português) foram feitas no programa ELAN (EUDICO *Linguistic Annotator*, 2020) e no Word. Também foram organizados léxicos para ambas as línguas e interlinearizados vários textos em Paresí no programa FLE_x (*FieldWorks Language Explorer*, 2019) para a análise presente na gramática de referência. As transcrições e traduções dos textos foram feitas pelos próprios falantes. A maioria das transcrições foi feita durante o período em que a pesquisadora principal não estava nas comunidades e, no caso do Paresí, foram revisadas posteriormente com os falantes. Os falantes de Paresí receberam treinamento sobre uso do programa ELAN, enquanto que os falantes de Enawenê-Nawê tiveram pouco contato com o programa e preferiram transcrever no Word usando o computador ou celular. Ainda estamos organizando as anotações das gravações dos Enawenê-Nawê para que todo material esteja no ELAN e possa posteriormente ser publicado em acervo digital.

Os arquivos no ELAN possuem três linhas de anotações: a da transcrição, a da tradução e a de notas. As informações mais detalhadas sobre glosas dos morfemas do Paresí ou do Enawenê-Nawê, chamada de *interlinearização*, estão organizadas no programa FLE_x, pois o mesmo permite a inserção automática das glosas (que já estão na base de dados), diferentemente do ELAN. Nenhuma das línguas possui uma ortografia estabelecida e usada consistentemente nas escolas indígenas. Por isso, as transcrições foram feitas nas ortografias que os falantes conhecem e, no caso do Paresí, foram sistematizadas para a ortografia que Brandão propôs para a língua.

Os metadados foram organizados inicialmente em planilha no Excel e depois inseridos no programa Lameta (HALTON et al., 2021) para que fossem criados arquivos IMDI (ISLE *Metadata Initiative*) para cada sessão. Cada uma dessas sessões possui o arquivo de áudio, vídeo e anotações no ELAN (formato .eaf) ou elicitaciones em PDF. No arquivo IMDI estão informações sobre as gravações, tais como: atores (pessoas envolvidas), assunto, descrição do conteúdo e palavras-

chave. Também há sessões com fotos organizadas nas seguintes categorias: comunidade, artesanato, escola, cotidiano, pesquisadores, treinamento, pessoas e trabalho.

Todos os dados do projeto Paresí foram armazenados no acervo do Museu Paraense Emílio Goeldi (MPEG), mas ainda não estão disponíveis para o público. Também foram armazenadas 20 horas de material coletado durante o período do apoio financeiro do ELDP (2011-2012) no acervo do *Endangered Language Archive* (ELAR)¹⁷; adicionalmente, 30 horas foram armazenadas no *Archive of the Indigenous Languages of the Americas* (AILLA)¹⁸, situado na Universidade de Texas em Austin, com alguns arquivos disponíveis on-line. Parte do material coletado no projeto Enawenê-Nawê foi organizado no acervo do ELAR¹⁹ e será disponibilizado também no acervo do MPEG.

Os acervos citados acima possuem uma classificação de códigos de acesso, pois alguns arquivos possuem acesso restrito. O ELAR possui os seguintes códigos: a) O para acesso livre, b) U para materiais que podem ser acessados através da criação de conta no acervo; e c) S para materiais que são de acesso restrito, e que só podem ser acessados com permissão do depositante. Estão marcados com o código de acesso S arquivos que podem conter dados pessoais dos falantes, tais como conversas e histórias de vida ou que ainda estão sendo analisados pelo depositante. Os materiais no acervo de AILLA possuem níveis de acesso parecidos, mas usando números no lugar de letras. Alguns arquivos exigem a criação de uma conta em seus sites antes que qualquer material seja acessado.

Esses acervos estão disponíveis em inglês (o ELAR) ou em inglês e espanhol (o AILLA), porém ainda não possuem informações de acesso em português, o que torna mais difícil para os indígenas terem acesso²⁰. Futuramente, pretendemos disponibilizá-los em acervos digitais regionais próximos às comunidades, como por exemplo, nas universidades (como na Universidade Federal do Mato Grosso), faculdades interculturais e escolas indígenas. Dessa forma, os falantes poderão ter acesso físico ao servidor onde estão os materiais, bem como o acesso mais fácil através da internet.

Para concluir esta seção, os dois projetos de documentação apresentados possibilitaram a organização de acervos linguísticos dessas línguas que estão disponíveis tanto para a comunidade acadêmica quanto para os membros das comunidades indígenas. No atual contexto da pandemia de COVID-19, os acervos são muito úteis para os pesquisadores que estão impossibilitados de visitarem as comunidades indígenas. Acima de tudo, os acervos são uma forma de salvaguardar o conhecimento tradicional desses povos, pois, infelizmente, os indígenas mais idosos, que são conhecedores da cultura, estão sendo vítimas do vírus. Consideramos que é de extrema importância que se faça o depósito dos materiais de línguas que ainda não estão arquivados em um acervo. Além disso, o próximo passo é fazer com que esses acervos estejam cada vez mais acessíveis para os

17 Coleção Paresi-Haliti no ELAR: <http://hdl.handle.net/2196/00-0000-0000-000E-2C87-4>

18 Coleção Paresi-Haliti no AILLA: <https://ailla.utexas.org/islandora/object/ailla:254756>

19 Coleção Enawenê-Nawê no ELAR: <http://hdl.handle.net/2196/00-0000-0000-0012-D797-0>

20 No AILLA, uma interface em língua portuguesa do acervo está sendo planejada e estará disponível em breve.

povos indígenas no Brasil. Dessa forma, os professores indígenas poderão utilizar os materiais em sala de aula nas comunidades e em seus trabalhos acadêmicos para as faculdades indígenas.

6. CONCLUSÃO

O Brasil e suas regiões vizinhas abrigam alguns dos mais altos níveis de diversidade linguística no mundo. Cerca de 300 línguas indígenas são faladas na América do Sul, e mais da metade desse número no Brasil (MOORE, 2007; GALUCIO, MOORE e VAN DER VOORT, 2018, p. 195). Esse é um fato comemorado pela Década Internacional das Línguas Indígenas da UNESCO, cujo início coincide com a redação deste artigo. No entanto, essa riqueza de línguas representa apenas uma fração do número que deve ter existido na América do Sul antes do contato europeu; e os processos de mudança e perda linguística continuam até os dias atuais. Quase 80 das 300 línguas restantes estão agora criticamente ameaçadas (MOORE, 2007). A perda das línguas indígenas no Brasil continua apesar de iniciativas como o Decreto n. 7.387 em 2010, que instituiu o Inventário Nacional da Diversidade Linguística (INDL), um programa governamental para pesquisar línguas e declará-las como patrimônio cultural imaterial do Brasil (GALUCIO, MOORE e VAN DER VOORT, 2018). Ainda hoje, a grande maioria das línguas indígenas brasileiras e sul-americanas carece de descrição e documentação substanciais. Talvez metade não tenha sequer um registro mínimo em acervos.

Como este artigo buscou demonstrar, o arquivamento é um componente essencial da documentação linguística. É somente por meio do arquivamento e por meio de repositórios digitais bem mantidos, com infraestrutura adequada e comprometimento institucional, que os materiais de documentação são preservados de maneira confiável e disponibilizados a longo prazo. Felizmente, o número de grandes projetos de documentação e coleções de acervos vem crescendo rapidamente nas últimas décadas, como exemplificado pelos projetos Paresí e Enawênê-Nawê descritos na seção final deste artigo (ver também MOORE e GALUCIO, 2016). Mas nós, linguistas, devemos construir nossa participação e investimento em arquivamento se quisermos manter os resultados deste trabalho documental. Como foi discutido aqui, isso significa contribuir, apoiar e expandir as iniciativas de arquivamento em todos os níveis: iniciativas regionais e nacionais como o Museu Goeldi no Brasil, bem como acervos de âmbito internacional como AILLA. Significa também apoiar e expandir as possibilidades para comunidades de falantes e seus descendentes acessarem esses materiais, tanto por meio de acervos, quanto em colaboração com estes. Como linguistas, temos a responsabilidade social e acadêmica de arquivar os materiais documentais que produzimos, apoiar iniciativas para arquivar materiais de legado, tornar os materiais tão acessíveis quanto possível e trabalhar em estreita colaboração com as comunidades para garantir um processo ético.

A pandemia do COVID-19 destaca a urgência desses esforços. Tragicamente, as comunidades estão perdendo anciãos e as línguas estão perdendo falantes. Os acervos de documentação linguística ajudam a preservar seu conhecimento para as próximas gerações e o tornam acessível a

membros das comunidades indígenas, acadêmicos e outros no futuro. Os acervos abrem novas possibilidades de pesquisa quando o trabalho de campo é impossível, oferecendo fontes alternativas de dados e análises e novos caminhos para a investigação. E os acervos podem fornecer recursos importantes para as comunidades que desejam revitalizar, manter ou simplesmente relembrar sua herança linguística e cultural.

AGRADECIMENTOS

Agradecemos aos organizadores da Abralín ao Vivo pela oportunidade de participar no seu fórum on-line (11 de julho de 2020). Gostaríamos também de agradecer aos membros das comunidades indígenas que participaram dos diversos projetos aqui citados e que estão representados nos acervos que mencionamos. Embora este estudo seja uma visão geral e não tenha recebido financiamento específico do projeto, agradecemos às entidades que generosamente apoiaram o trabalho que se reflete nas coleções de acervo aqui mencionadas. O AILLA agradece o financiamento contínuo da *National Science Foundation* e do *National Endowment for the Humanities* ao longo de muitos anos. O CLA agradece o financiamento recente do *National Endowment for the Humanities* (PD-271354-20; 2020-2023) e o atual financiamento da organização Arcadia. O Museu Goeldi agradece pelo apoio dos programas ELDP e DoBeS, da USAID, da UNESCO, do Fundo de Embaixador da Embaixada dos EUA, do Ministério de Ciência, Tecnologia e Inovação (MCTI), do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e do Fundo de Defesa de Direitos Difusos (FDD) do Ministério da Justiça e Segurança Pública. Por fim, gostaríamos de agradecer a Thainá de Lima Reis pela ajuda na tradução e formatação.

INFORMAÇÕES COMPLEMENTARES

CONFLITO DE INTERESSE

Os autores não têm conflitos de interesse a declarar.

DECLARAÇÃO DE DISPONIBILIDADE DE DADOS

Este artigo representa um estudo geral sem dados primários específicos do projeto. Links de identificador persistente para conteúdo relevante dos acervos foram incluídos ao longo do texto, quando relevantes.

FONTES DE FINANCIAMENTO

O projeto de doutorado de Rosés Labrada, relatado na Seção 3, foi financiado pelo Conselho de Pesquisa em Ciências Sociais e Humanas do Canadá (SSHRC) por meio de uma Bolsa de Pós-Graduação Vanier Canada [Prêmio No. 770-2012-0151]. Os trabalhos de Ana Paula Brandão sobre Paresí-Haliti e Enenawê-Nawê, relatados na Seção 5, foram financiados pela *National Science Foundation* (BCS 1123943) e pelo *Endangered Languages Documentation Program* (IGSO160, IPF0280).

AVALIAÇÃO E RESPOSTA DOS AUTORES

Avaliação: <https://doi.org/10.25189/2675-4916.2023.V4.N1.ID666.R>

Resposta dos Autores: <https://doi.org/10.25189/2675-4916.2023.V4.N1.ID666.A>

REFERÊNCIAS

AIKHENVALD, Alexandra. The Arawak language family. In: R. M. W. Dixon & A. Y. Aikhenvald (orgs.). *The Amazonian languages*. Cambridge: Cambridge University Press, 1999, p. 65-106.

ANWAR, Yasmin. Chance phone call keeps alive scholar's remarkable Amazonian legacy. In: *Berkeley News*, May 24, 2021. Disponível em: <https://news.berkeley.edu/2021/05/24/chance-phone-call-keeps-alive-scholars-remarkable-amazonian-legacy/>.

Assembleia Geral das Nações Unidas (AGNU). *Declaração das Nações Unidas sobre os Direitos dos Povos Indígenas*, 2 out. 2007, A/RES/61/295. Disponível em: https://www.un.org/esa/socdev/unpfii/documents/DRIPS_en.pdf (inglês), https://www.un.org/esa/socdev/unpfii/documents/DRIPS_pt.pdf (português).

AUSTIN, Peter K. Language documentation & legacy text materials. *Asian and African Languages and Linguistics*, n. 11, p. 23-44, 2017.

BALDWIN, Daryl; PÉREZ BÁEZ, Gabriela; HINTON, Leanne. The Breath of Life workshops and institutes. In: HINTON, Leanne; HUSS, Leena; ROCHE, Gerald (orgs.). *The Routledge handbook of language revitalization*. London: Routledge, 2018, p. 188-196.

BALLARD, Sarah; APPLGATE, Richard B. *Bodega Miwok sound recordings 2014-05*, California Language Archive, Survey of California and Other Indian Languages, University of California, Berkeley, 1974. Disponível em: <http://dx.doi.org/doi:10.7297/X2QN64XN>.

BALLARD, Sarah; CALLAGHAN, Catherine A. *The Catherine A. Callaghan collection of Bodega Miwok sound recordings LA 6*, California Language Archive, Survey of California and Other Indian Languages, University of California, Berkeley, 1960. Disponível em: <http://cla.berkeley.edu/collection/10086>.

BARWICK, Linda. Turning it all upside down...Imagining a distributed digital audiovisual archive. *Literary and Linguistic Computing*, v. 19, n. 3, p. 253-263, 2004. Disponível em: <https://doi.org/10.1093/llc/19.3.253>.

BARWICK, Linda; MARETT, Allan; WALSH, Michael; REID, Nicholas; FORD, Lysbeth. Communities of interest: Issues in establishing a digital resource on Murrinh-Patha song at Wadeye (Port Keats), NT. *Literary and Linguistic Computing*, v. 20, n. 4, p. 383-397, 2005. Disponível em: <https://doi.org/10.1093/llc/fqi048>

BEREZ-KROEKER, Andrea L.; GAWNE, Lauren; KUNG, Susan Smythe; KELLY, Barbara F.; HESTON, Tyler; HOLTON, Gary; PULSIFER, Peter; BEAVER, David I.; CHELLIAH, Shobhana; DUBINSKY, Stanley; MEIER, Richard P.; THIEBERGER, Nick; RICE, Keren; WOODBURY, Anthony C. Reproducible research in linguistics: A position statement on data citation and attribution in our field. *Linguistics*, v. 56, n. 1, p. 1-18, 2018. Disponível em: <https://doi.org/10.1515/ling-2017-0032>.

BEREZ-KROEKER, Andrea L.; HENKE, Ryan. Language archiving. In: REHG, Kenneth L.; CAMPBELL, Lyle (orgs.) *The Oxford handbook of endangered languages*. Oxford: Oxford University Press, 2018, p. 347-369. Disponível em: <https://doi.org/10.1093/oxfordhb/9780190610029.013.18>.

BISCHOFF, Shannon T.; JANY, Carmen (orgs.). *Insights from practices in community-based research: From theory to practice around the globe*. Berlin & Boston: De Gruyter Mouton, 2018. Disponível em: <https://doi.org/10.1515/9783110527018>.

BOMFIM, Anari Braz. Patxohã: A retomada da língua do povo Pataxó. *Revista Lingüística*, v. 13, n. 1, p. 303-327, 2017. Disponível em: <https://revistas.ufjr.br/index.php/rl>.

BRANDÃO, Ana Paula. *Verb morphology in Paresi-Haliti (Aruak)*. Qualifying paper, University of Texas at Austin, 2010.

BRANDÃO, Ana Paula. *A reference grammar of Paresi-haliti (Aruák)*. Tese (Doutorado em Linguística), University of Texas at Austin, 2014.

BRANDÃO, Ana Paula; CARVALHO, Fernando; PEREIRA, Everton. Estudo histórico-comparativo preliminar do subgrupo Juruena (Aruák). Trabalho apresentado no *Congresso Internacional de Estudos Linguísticos e Literários na Amazônia* (VI CIELLA), Belém, 2018.

BRANDÃO, Ana Paula; REIS, Thainá de Lima. Gênero gramatical em Enawene Nawe? *Revista de Letras Norte@mento*. Dossiê Temático: Para a década das línguas indígenas, Sinop, v. 13, n. 33, p. 208-227, 2020.

BRANDÃO, Ana Paula; ZEZOKIWARE, Jurandir. A documentação participativa: O caso das línguas Paresi e Enawene Nawe. *Revista Moara*, n. 50, p. 87-106, 2018. Disponível em: <http://dx.doi.org/10.18542/moara.v1i50.6815>.

CARROLL, Stephanie Russo; GARBA, Ibrahim; FIGUEROA-RODRÍGUEZ, Oscar L.; HOLBROOK, Jarita; LOVETT, Raymond; MATERECHERA, Simeon; PARSONS, Mark; RASEROKA, Kay; RODRIGUEZ-LONEBEAR, Desi; ROWE, Robyn; SARA, Rodrigo; WALKER, Jennifer D.; ANDERSON, Jane; HUDSON, Maui. The CARE principles for Indigenous data governance. *Data Science Journal*, v. 19, n. 1, p.1-12, 2020. Disponível em: <https://doi.org/10.5334/dsj-2019-031>.

CONATHAN, Lisa. Archiving and language documentation. In: AUSTIN, Peter K.; SALLABANK, Julia (orgs.). *The Cambridge handbook of endangered languages*. Cambridge: Cambridge University Press, 2011, p. 235-254. Disponível em: <https://doi.org/10.1017/CBO9780511975981.012>.

CONZETT, Philipp; DE SMEDT, Koenradd. Guidance for citing linguistic data. In: BEREZ-KROEKER, Andrea L.; MCDONNELL, Bradley; KOLLER, Eve; COLLISTER, Lauren B. (orgs.). *The open handbook of linguistic data management*. Cambridge: The MIT Press, 2022, p. 143-155. Disponível em: <https://doi.org/10.7551/mitpress/12200.001.0001>.

COWELL, Andrew. *Time-aligned annotations of Bodega Miwok sound recordings 2019-18*. California Language Archive, Survey of California and Other Indian Languages, University of California, Berkeley, 2019+. Disponível em: <http://dx.doi.org/doi:10.7297/X2251GCO>.

CZAYKOWSKA-HIGGINS, Ewa. Research models, community engagement, and linguistic fieldwork: Reflections on working within Canadian Indigenous communities. *Language Documentation & Conservation*, v. 3, n. 1, p. 15-50, 2009.

DEBENPORT, Erin. The potential complexity of 'universal ownership': Cultural property, textual circulation, and linguistic fieldwork. *Language & Communication*, v. 30, n. 3, p. 204-210, 2010.

DIGITAL PRESERVATION COALITION (DPC). Glossary. *Digital preservation handbook*. Ed. 2. 2015. Disponível em: <https://www.dpconline.org/handbook/glossary>

DOBRIN, Lise M. From linguistic elicitation to eliciting the linguist: Lessons in community empowerment from Melanesia. *Language*, v. 84, n. 2, p. 300-324, 2008.

DOBRIN, Lise M.; SCHWARTZ, Saul. The social lives of linguistic legacy materials. *Language Documentation and Description*, v. 21, p. 1-36, 2021.

DWYER, Arienne M. Ethics and practicalities of cooperative fieldwork and analysis. In: GIPPERT, Jost; HIMMELMANN, Nikolaus P.; MOSEL, Ulrike (orgs.). *Essentials of language documentation*. Berlin: Mouton de Gruyter, 2006, p. 31-66.

EUDICO Linguistic Annotator (ELAN) (version 5.9). Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Disponível em: <https://tla.mpi.nl/tools/tla-tools/elan/>. 2020.

FABRE, Alain. *Diccionario etnolingüístico y guía bibliográfica de los pueblos indígenas sudamericanos, Arawak*. 2005. Disponível em: <http://www.ling.fi/Entradas%20diccionario/Dic=Arawak.pdf>.

FieldWorks Language Explorer (FLEX) (version 8.3.12). SIL. 2019. Disponível em: <https://software.sil.org/fieldworks/>

GALUCIO, Ana Vilacy; MOORE, Denny; VAN DER VOORT, Hein. O Patrimônio linguístico do Brasil: Novas perspectivas e abordagens no planejamento e gestão de uma política de diversidade linguística. *Revista do Patrimônio Histórico e Artístico Nacional*, n. 38, p. 194-219, 2018.

GASSER, Emily. The right to say yes: Language documentation in West Papua. *Australian Journal of Linguistics*, v. 37, n. 4, p. 502-526, 2017. Disponível em: <https://doi.org/10.1080/07268602.2017.1350131>.

GEHR, Susan. *Breath of Life: Revitalizing California's native languages through archives*. Dissertação (Mestrado em Linguística), San José State University, 2013. <https://doi.org/10.31979/etd.yfva-e77q>.

GEORGE, Christine; JOSEPH, Lawrence; XAVIER, Felix; NOBLE JR, Gladwyn K. *Linguistic materials on indigenous languages of Guyana 2018-03*, California Language Archive, Survey of California and Other Indian Languages, University of California, Berkeley, 1965 Disponível em: <http://dx.doi.org/doi:10.7297/X2NC5ZCN>.

GLENN, Akiemi. Five dimensions of collaboration: Toward a critical theory of coordination and interoperability in language documentation. *Language Documentation & Conservation*, v. 3, n. 2, p. 149-160, 2009.

GODDARD, Ives. Philological approaches to the study of North American Indian languages: Documents and documentation. In: SEBEOK, Thomas A. (org.). *Current trends in Linguistics: Linguistics in North America*. vol. 10. The Hague: Mouton, 1973, p. 727-745.

GOOD, Jeff. The scope of linguistic data. In: BEREZ-KROEKER, Andrea L.; MCDONNELL, Bradley; KOLLER, Eve; COLLISTER, Lauren B. (orgs.). *The open handbook of linguistic data management*. Cambridge, MA: The MIT Press, 2022, p. 27-47.

GRINEVALD, Colette. Worrying about ethics and wondering about 'informed consent': Fieldwork from an Americanist perspective. In: SAXENA, Anju; BORIN, Lars (orgs.). *Lesser-known languages of South Asia: Status and policies, case studies and applications of information technology*. Berlin: De Gruyter Mouton, 2006, p. 339-370. Disponível em: <https://doi-org/10.1515/9783110197785.3.339>.

HALTON, John; HOLTON, Gary; SEYFEDDINIPUR, Mandana; THIEBERGER, Nicholas. *Lameta [software]*. 2021. Disponível em: <https://github.com/onset/laMETA/releases>.

HAN, Na-Rae. Transforming data. In: BEREZ-KROEKER Andrea L.; MCDONNELL, Bradley; KOLLER, Eve; COLLISTER, Lauren B. (orgs.) *The open handbook of linguistic data management*. Cambridge: The MIT Press, 2022, p. 73-87. Disponível em: <https://doi.org/10.7551/mitpress/12200.001.0001>.

HENKE, Ryan; BEREZ-KROEKER, Andrea L. A brief history of archiving in language documentation, with an annotated bibliography. *Language Documentation & Conservation*, v. 10, p. 411-457, 2016.

HIMMELMANN, Nikolaus P. Documentary and descriptive linguistics. *Linguistics*, v. 36, p. 161-195, 1998.

HIMMELMANN, Nikolaus P. Language documentation: What is it and what is it good for? In: GIPPERT, Jost; HIMMELMANN, Nikolaus P.; MOSEL, Ulrike (orgs.). *Essentials of language documentation*. Berlin: Mouton de Gruyter, 2006, p. 1-30.

HIMMELMANN, Nikolaus P. Linguistic data types and the interface between language documentation and description. *Language Documentation & Conservation*, v. 6, p. 187-207, 2012.

HOLTON, Gary. Relatively ethical: A comparison of linguistic research paradigms in Alaska and Indonesia. *Language Documentation & Conservation*, v. 3, n. 2, p. 161-175, 2009.

HOLTON, Gary; LEONARD, Wesley Y.; PULSIFER, Peter L. Indigenous Peoples, Ethics, and Linguistic Data. In: BEREZ-KROEKER Andrea L.; MCDONNELL, Bradley; KOLLER, Eve; COLLISTER, Lauren B. (orgs.). *The open handbook of linguistic data management*. Cambridge: The MIT Press, 2022, p. 49-60. Disponível em: <https://doi.org/10.7551/mitpress/12200.001.0001>.

INNES, Pamela. Ethical problems in archival research: Beyond accessibility. *Language & Communication*, v. 30, n. 3, p. 198-203, 2010. Disponível em: <https://doi.org/10.1016/j.langcom.2009.11.006>.

JOHNSON, Heidi. Language documentation and archiving, or how to build a better corpus. *Language Documentation and Description*, v. 2, p. 140-153, 2004.

KAPLAN, Judith; LEMOV, Rebecca. Archiving endangerment, endangered archives: Journeys through the sound archives of Americanist anthropology and linguistics, 1911-2016. *Technology and Culture*, v. 60, n. 2S, p. S161-S187, 2019. Disponível em: <https://doi.org/10.1353/tech.2019.0067>.

KOENIG, A.; RINGERSMA, J.; TRILSBEEK, P. The Language Archiving Technology domain. In: VETULANI, Z. (org.), *Human Language Technologies as a Challenge for Computer Science and Linguistics*. Nijmegen: MPI Psycholinguistics, 2009, p. 295-299.

KUNG, Susan Smythe. Data archiving, access, and repatriation. In: STANLAW, James (org.). *The international encyclopedia of linguistic anthropology*. Wiley On-line Library, 2020. Disponível em: <https://doi.org/10.1002/9781118786093.iela0430>.

KUNG, Susan Smythe; SULLIVANT, Ryan; POJMAN, Elena; NIWAGABA; Alicia. *Archiving for the future: Simple steps for archiving language documentation collections [OER]*. 2020. Disponível em: <https://archivingforthefuture.teachable.com/>.

LEONARD, Wesley Y.; HAYNES, Erin. Making 'collaboration' collaborative: An examination of perspectives that frame linguistic field research. *Language Documentation & Conservation*, v. 4, p. 268-293, 2010.

LUKANIEC, Megan. Managing data from archival documentation for language reclamation. In: BEREZ-KROEKER, Andrea L.; MCDONNELL, Bradley; KOLLER, Eve; COLLISTER, Lauren B. (orgs.). *The open handbook of linguistic data management*. Cambridge: The MIT Press, 2022, p. 315-325. Disponível em: <https://doi.org/10.7551/mitpress/12200.001.0001>.

MACRI, Martha J.; SARMENTO; James. Respecting privacy: Ethical and pragmatic considerations. *Language & Communication*, v. 30, n. 3, p. 92-97, 2010. Disponível em: <https://doi.org/10.1016/j.langcom.2009.11.005>.

MILLER, Ellis R. *Phylogenetic classification of the Negro-Roraima subgroup*. Trabalho de Conclusão de Curso (BA), University of California, Berkeley, 2021.

MILLER, Robert J. Introduction. In: LINK, Adrianna; SHELTON, Abigail; SPERO, Patrick (orgs.). *Indigenous languages and the promise of archives*. Lincoln: University of Nebraska Press, 2021, p.1-24.

MOORE, Denny. Endangered languages of lowland tropical South America. In: BREZINGER, Matthias (org.). *Language diversity endangered*. Berlin: Mouton de Gruyter, 2007, p. 29-58.

MOORE, Denny; GALUCIO, Ana Vilacy. Perspectives for the documentation of indigenous languages in Brazil. In: PÉREZ BÁEZ, Gabriela; ROGERS, Chris; ROSÉS LABRADA, Jorge Emilio (orgs.). *Language documentation and revitalization in Latin American contexts*. Berlin & Boston: De Gruyter Mouton, 2016, p. 29-58.

MOORE, Denny; GALUCIO, Ana Vilacy; GABAS JR, Nilson. Desafio de documentar e preservar as línguas Amazônicas. *Scientific American Brasil*, n. 3, p. 36-43, 2008.

NATHAN, David. *Progressive archiving: Theoretical and practical implications for documentary linguistics*. Presentation at the International Conference on Language Documentation and Conservation, March 3, 2013. Disponível em: <http://hdl.handle.net/10125/26115>.

NATHAN, David; AUSTIN, Peter K. Reconceiving metadata: Language documentation through thick and thin. *Language Documentation and Description*, v. 2, p. 179-189, 2004.

O'HAGAN, Zachary. *A phonological sketch of Atorai (Arawak, Guyana) based on unique recordings*. Presentation at Fieldwork Forum, Berkeley, May 2 2018.

O'HAGAN, Zachary. Obituario, Gerald Weiss (1932-2021). *Amazonía peruana*, n. 34, p. 279-286, 2021.

O'MEARA, Carolyn; GONZÁLEZ GUADARRAMA, Octavio Alonso. Accessibility to results and primary data of research on indigenous languages of Mexico. In: PÉREZ BÁEZ, Gabriela; ROGERS, Chris; ROSÉS LABRADA, Jorge Emilio (orgs.). *Language documentation and revitalization in Latin American contexts*. Berlin & Boston: De Gruyter Mouton, 2016, p. 59-80. Disponível em: <https://doi.org/10.1515/9783110428902>.

O'MEARA, Carolyn; GOOD, Jeff. Ethical issues in legacy language resources. *Language & Communication*, v. 30, n. 3, p. 162-170, 2010.

OWENS, Trevor. *The theory and craft of digital preservation*. Baltimore: John Hopkins University Press, 2018.

PARKER, Verdona. *The Verdona Parker collection of Hupa sound recordings and films LA 256*. California Language Archive, Survey of California and Other Indian Languages, University of California, Berkeley, 2003+. Disponível em: <http://dx.doi.org/doi:10.7297/X29SIPBD>.

PARKER, Verdona; CAMPBELL, Amy; ESCAMILLA, Ramón; NEWBOLD, Lindsey; SPENCE, Justin. *Materials of the Hupa Language Documentation Project 2017-06*. California Language Archive, Survey of California and Other Indian Languages, University of California, Berkeley, 2005+. Disponível em: <http://dx.doi.org/doi:10.7297/X22R3Q2G>.

PAYNE, David L. A classification of Maipuran (Arawakan) languages based on shared lexical retentions. In: DERBYSHIRE, Desmond; PULLUM, Geoffrey (orgs.). *Handbook of Amazonian languages*. vol. 3. Berlin: Mouton de Gruyter, 1991, p. 355-499.

PEREIRA, Everton. *Estudo histórico-comparativo preliminar das línguas Paresi e Enawene Nawe*. Trabalho de Conclusão de Curso, Universidade Federal do Pará, 2018.

PÉREZ BÁEZ, Gabriela (fotógrafa). *Kaufman collection acquisition trip photos (DSC_0001.JPG)*. Archiving the Terrence Kaufman collection. The Archive of the Indigenous Languages of Latin America, ailla.utexas.org. Access: Public. PID ailla:257492.

PÉREZ BÁEZ, Gabriela (fotógrafa). *Kaufman collection acquisition trip photos (DSC_0052.JPG)*. Archiving the Terrence Kaufman collection. The Archive of the Indigenous Languages of Latin America, ailla.utexas.org. Access: Public. PID ailla:257543.

PÉREZ BÁEZ, Gabriela; ROGERS, Chris; ROSÉS LABRADA, Jorge Emilio. *Language documentation and revitalization in Latin American contexts*. Berlin & Boston: De Gruyter Mouton, 2016. Disponível em: <https://doi.org/10.1515/9783110428902>.

RAMIREZ, Henri. *Línguas Arawak da Amazônia setentrional: Comparação e descrição*. Manaus: Editora da Universidade do Amazonas, 2001.

REIS, Thainá de Lima. *Uma análise preliminar do gênero gramatical em Enawene Nawe (Aruák)*. Trabalho de Conclusão de Curso, Universidade Federal do Pará, 2020.

REZENDE, Ubiray. *Fonética e fonologia da língua Enawene-Nawe (Aruak): Uma primeira abordagem*. Dissertação (Mestrado em Linguística), Universidade Federal do Rio de Janeiro, 2003.

REZENDE, Ubiray. *Aspectos da gramática da língua Enawene-Nawe (Aruak)*. Tese (Doutorado em Linguística), Universidade Federal do Rio de Janeiro, 2013.

RICE, Keren. Ethical issues in linguistic fieldwork: An overview. *Journal of Academic Ethics*, v. 4, p. 123-155, 2006.

RICE, Alexander. Using YouTube as the primary transcription and translation platform for remote corpus work. *Language Documentation & Conservation*, v. 15, p. 514-550, 2021.

ROBINSON, Laura C. Archiving directly from the field. In: BARWICK, Linda; THIEBERGER, Nicholas (orgs.). *Sustainable data from digital fieldwork*. Sydney: University of Sydney Press, 2006, p. 23-32.

ROBINSON, Laura C. Informed consent among analog people in a digital world. *Language & Communication*, v. 30, n. 3, p. 186-191, 2010. Disponível em: <https://doi.org/10.1016/j.langcom.2009.11.002>.

ROSÉS LABRADA, Jorge Emilio. *The Mako language: Vitality, grammar and classification*. Tese (Doutorado em Estudos Franceses e em Ciências da Linguagem), University of Western Ontario & Université Lumière-Lyon, 2015. Disponível em: <https://ir.lib.uwo.ca/etd/2851>.

SANTOS, Wesley dos. *Kawahiva language documentation archive 2019-06*. California Language Archive, Survey of California and Other Indian Languages, University of California, Berkeley, 2017+. Disponível em: <http://dx.doi.org/doi:10.7297/X2P26W9H>.

SAPIÉN, Racquel-María. Design and implementation of collaborative language documentation projects. In: REHG, Kenneth L.; CAMPBELL, Lyle (orgs.). *The Oxford handbook of endangered languages*. Oxford: Oxford University Press, 2018, p. 203-224.

SEIFART, Frank; DRUDE, Sebastian; FRANCHETTO, Bruna; GASCHÉ, Jürg; GOLLUSCIO, Lucía; MANRIQUE, Elizabeth. Language documentation and archives in South America. *Language Documentation & Conservation*, vol. 2, n. 1, p. 130-140, 2008.

SILVA, Glauber Romling da. *Fonologia da língua Paresi-Haliti (Aruak)*. Dissertação (Mestrado em Linguística), Universidade Federal do Rio de Janeiro, 2009.

SILVA, Glauber Romling da. *Morfossintaxe da língua Paresi-Haliti*. Tese (Doutorado em Linguística), Universidade Federal do Rio de Janeiro, 2013.

SPENCE, Justin. Learning languages through archives. In: HINTON, Leanne; HUSS, Leena; ROCHE, Gerald (orgs.). *The Routledge handbook of language revitalization*. New York: Routledge, 2018, p. 179-187.

THIEBERGER, Nicholas. Lost and found: Linguists and musicologists at three Australian universities are working together to preserve rare recordings and make them accessible to communities across the Pacific and beyond. *The ACU Review*, 2019. Disponível em: <https://www.acu.ac.uk/the-acu-review/lost-and-found/>.

THIEBERGER, Nicholas; BEREZ, Andrea L. Linguistic data management. In: THIEBERGER, Nicholas (org.). *The Oxford handbook of linguistic fieldwork*. Oxford: Oxford University Press, 2012, p. 90-118.

VAPNARSKY, Valentina. Circulation et virtualités des savoirs amérindiens à l'ère du numérique – From home base to database...and back? The circulation and virtualities of Amerindian knowledge in the digital era. *Journal de la Société des Américanistes*, v. 106, n. 2, p. 79-104, 2020.

WEISS, Gerald. *The cosmology of the Campa Indians of Eastern Peru*. PhD dissertation, University of Michigan, 1969.

WILBUR, Joshua. Archiving for the community: Engaging local Archives in language documentation projects. *Language Documentation and Description*, v. 12: Special Issue on Language Documentation and Archiving, p. 85-102, 2014.

WOODBURY, Anthony C. Defining documentary linguistics. *Language Documentation and Description*, v. 1, p. 35-51, 2003.

YAMADA, Racquel-Maria. Collaborative linguistic fieldwork: Practical application of the Empowerment Model. *Language Documentation & Conservation*, v. 1, n. 2, p. 257-282, 2007.