



UNIVERSIDADE FEDERAL  
DO RIO DE JANEIRO



**UNIVERSIDADE FEDERAL DO RIO DE JANEIRO**  
**MUSEU NACIONAL**

**MESTRADO PROFISSIONAL EM LINGUÍSTICA**  
**E LÍNGUAS INDÍGENAS**

**O IMPACTO DOS PROGRAMAS DOBES, DMA E AILLA**  
**SOBRE A QUESTÃO DAS LÍNGUAS INDÍGENAS**  
**AMEAÇADAS: UMA ABORDAGEM CRÍTICA E PROPOSTA**  
**BÁSICA DE UM PROJETO DE ARQUIVO LINGUÍSTICO**  
**PARA O SETOR DE LINGUÍSTICA DO MUSEU NACIONAL**

**Luiz Claudio Falcão de Albuquerque**

Rio de Janeiro

2023



UNIVERSIDADE FEDERAL  
DO RIO DE JANEIRO



**UNIVERSIDADE FEDERAL DO RIO DE JANEIRO**  
**MUSEU NACIONAL**

**MESTRADO PROFISSIONAL EM LINGUÍSTICA**  
**E LÍNGUAS INDÍGENAS**

**O IMPACTO DOS PROGRAMAS DOBES, DMA E AILLA**  
**SOBRE A QUESTÃO DAS LÍNGUAS INDÍGENAS**  
**AMEAÇADAS: UMA ABORDAGEM CRÍTICA E PROPOSTA**  
**BÁSICA DE UM PROJETO DE ARQUIVO LINGUÍSTICO**  
**PARA O SETOR DE LINGUÍSTICA DO MUSEU NACIONAL**

**Luiz Claudio Falcão de Albuquerque**

Dissertação de Mestrado apresentada ao Mestrado Profissional em Linguística e Línguas Indígenas da Universidade Federal do Rio de Janeiro como requisito para a obtenção do Título de Mestre em Linguística e Línguas Indígenas.

**Orientadora: Profa. Doutora Marília Lopes da Costa Facó Soares (UFRJ)**  
**Linha de pesquisa: Descrição, Análise e Documentação de Línguas Indígenas**

Rio de Janeiro  
2023

A345i Albuquerque, Luiz Claudio Falcão de  
O impacto dos programas DOBES, DMA E AILLA sobre a questão das línguas indígenas ameaçadas: uma abordagem crítica e proposta básica de um projeto de arquivo linguístico para o Setor de Linguística do Museu Nacional / Luiz Claudio Falcão de Albuquerque. – Rio de Janeiro, 2023.

225f. : il. (color.)

Orientadora: Marília Lopes da Costa Facó Soares

Dissertação (mestrado) - Universidade Federal do Rio de Janeiro, Museu Nacional, Mestrado Profissional em Linguística e Línguas Indígenas - PROFLLIND, 2023.

1.Línguas indígenas. 2. Programas DOBES, DMA e AILLA. 3. Projeto de Arquivo Linguístico. I. Soares, Marília Lopes da Costa Facó. III.Título.

**CDD498**



UNIVERSIDADE FEDERAL  
DO RIO DE JANEIRO



**UNIVERSIDADE FEDERAL DO RIO DE JANEIRO  
MUSEU NACIONAL**

**MESTRADO PROFISSIONAL EM LINGUÍSTICA  
E LÍNGUAS INDÍGENAS**

**O IMPACTO DOS PROGRAMAS DOBES, DMA E AILLA  
SOBRE A QUESTÃO DAS LÍNGUAS INDÍGENAS  
AMEAÇADAS: UMA ABORDAGEM CRÍTICA E PROPOSTA  
BÁSICA DE UM PROJETO DE ARQUIVO LINGUÍSTICO  
PARA O SETOR DE LINGUÍSTICA DO MUSEU NACIONAL**

**Luiz Claudio Falcão de Albuquerque**

**Banca Examinadora**

---

**Presidente: Dra. Marília Lopes da Costa Facó Soares (PROFLLIND-  
MN-UFRJ)**

---

**Prof. Dr. Fernando Orphão de Carvalho (PROFLLIND- MN- UFRJ)**

---

**Profa. Dra. Christina Abreu Gomes (Pós-Ling-UFRJ)**

---

**Prof. Dr. Gean Nunes Damulakis (PROFLLIND- MN- UFRJ)  
(suplente)**

---

**Prof. Dr. Marcelo Alexandre Lopes de Melo (Pós-Ling-UFRJ)  
(suplente)**

Música de trabalho

***Sem trabalho eu não sou nada  
Não tenho dignidade  
Não sinto o meu valor  
Não tenho identidade  
Mas o que eu tenho  
É só um emprego  
E um salário miserável  
Eu tenho o meu ofício  
Que me cansa de verdade  
Tem gente que não tem nada  
E outros que têm mais do que precisam  
Tem gente que não quer saber de trabalhar  
...  
E quando chega o fim do dia  
Eu só penso em descansar  
E voltar p'rá casa, pros teus braços  
Quem sabe esquecer um pouco  
Do pouco que não temos  
Quem sabe esquecer um pouco  
De tudo que não sabemos***

(Marcelo Bonfá / Dado Villa-Lobos / Renato Russo)

Ao meu querido pai, **Sebastião** (*in memoriam*), um retirante nordestino. Homem que chegou ao sudeste em busca de melhores condições de vida ainda analfabeto. Apesar das dificuldades, formou a própria família, e criou, sozinho, os seis filhos sem jamais lhes negar a valorosa educação pública e a leitura de livros.

À minha madrasta, **Otília** (*in memoriam*), pessoa que tratou a mim e a meus irmãos como uma verdadeira mãe. Prova de sua grandeza: certo dia pegou nas mãos de um menino, de sete anos, e o levou à escola pela primeira vez, contrariando os tempos difíceis. Desde então, ele nunca mais deixou de estudar.

Aos meus fraternos **irmãos**, pessoas que apesar da distância, atravessam comigo essa imprevisível e curiosa jornada existencial.

Aos tios **Benedito** e **Maria José**, figuras admiráveis que inspiraram a meus **primos** e a mim a descobrir a vida com alegria, coragem e curiosidade.

Às pessoas mais importantes na minha vida, minha esposa **Elcimara** e minhas três filhas: **Mayra**, **Gabriela Luiza** (*in memoriam*) e **Manuela**. Se ainda resisto às adversidades e angústias cotidianas, a origem de minha resiliência está na existência da minha família e no amor que compartilhamos nesta e noutras possíveis formas de vida.

## AGRADECIMENTOS

À orientadora, professora Dra. Marília Lopes da Costa Facó Soares, gostaria de agradecer por tornar possível o desenvolvimento desta dissertação. Desde as fases iniciais, a professora anteviu que meu projeto poderia se transformar em material importante para a elaboração de material acadêmico-profissional destinado à criação de um arquivo linguístico digital para o Setor de Linguística do Museu Nacional. Nesse sentido, foi fundamental a sua orientação para a definição do tema, a indicação das referências bibliográficas, o apontamento dos ajustes no texto e as críticas, construtivas e necessárias, ao longo da redação do trabalho.

À Universidade Federal do Rio de Janeiro (UFRJ) ao permitir, através do Programa de Qualificação Institucional (PQI/UFRJ), que o pessoal docente e os técnico-administrativos em Educação tivessem acesso aos Programas de Pós-graduação no nível *stricto sensu* (mestrado e doutorado).

Ao corpo docente do Mestrado Profissional em Linguística e Línguas Indígenas do Museu Nacional/UFRJ pela qualidade das aulas, o compromisso com a educação pública de qualidade e a dedicação e ajuda nos momentos necessários. Aproveito e presto um singelo agradecimento (*in memoriam*) à querida e saudosa professora Márcia Damaso Vieira de quem guardo as lembranças das excelentes aulas e do bom humor.

À bibliotecária do Mestrado Profissional em Linguística e Línguas Indígenas do Museu Nacional/UFRJ, Cristina, pela boa vontade na sugestão de textos e indicação dos locais onde poderia localizá-los para estudos e pesquisas.

Aos professores do SENAC-RJ onde fiz o curso de Web Design, no período de 2018-2019, e onde conservo algumas amizades. Esses profissionais foram importantes nas sugestões técnicas e informais sobre as etapas de desenvolvimento da proposta básica de projeto de arquivo linguístico para o Setor de Linguística do Museu Nacional, ao indicar os livros, artigos e vídeos essenciais sobre modelagem de sistemas, bancos de dados e prototipagem de telas de aplicações web.

Às instituições ligadas aos programas de documentação DOBES, DMA e AILLA por permitirem, através de seus canais de comunicação, resolver dúvidas e ter acesso a artigos e publicações diversas sobre seus acervos. Com isso, em alguns casos, atualmente, tenho o contato de pessoas ligadas à documentação linguística, entre as quais a professora Lenore Grenoble e o pesquisador Peter K. Austin, entre outros, dos quais recebo os informes de suas publicações mais recentes.

Por fim, gostaria de agradecer, mais uma vez, à minha família pela compreensão por tantos momentos de lazer não vividos e o encolhimento do convívio social entre nós mesmos e com as pessoas que mais amamos. Porém, como afirma o poeta português Fernando Pessoa: “Tudo vale a pena se a alma não é pequena”.

## RESUMO

Esta dissertação tem como objetivo geral investigar, criticamente, o impacto dos programas de documentação linguística Dokumentation Bedrohter Sprachen (DOBES), Digital Media Archive (DMA) e Archive of Indigenous Languages of Latin America (AILLA) a partir de seus processos de estruturação para enfrentar o problema relacionado às línguas indígenas ameaçadas. Para esse propósito, a ideia é avaliar, amplamente, de que forma o conjunto de iniciativas adotadas podem ser validadas ou não como políticas eficientes em torno dos interesses desejados. Em termos de referencial teórico, emprega-se material relativo às publicações mais recentes relacionadas ao nosso objeto de pesquisa. Acerca da metodologia, adota-se os elementos essenciais divididos conforme categorias de acordo com: a finalidade (pesquisa aplicada), a abordagem (pesquisa quali-quantitativa), o método (hipotético-dedutivo) e os procedimentos (pesquisa bibliográfica). Acerca do resultado daquilo que foi analisado sobre o objeto, pode-se afirmar que os programas de documentação linguística estudados atuam com profissionalismo, estão consolidados e atingiram seus objetivos em termos de documentação, armazenamento e disponibilização dos arquivos linguísticos ao público em geral. O impacto dos programas DOBES, DMA e AILLA sobre a questão das línguas indígenas ameaçadas é positivo, pois eles podem ser considerados relevantes no esforço de evitar-se uma catástrofe linguística e cultural irreparável para os povos indígenas, em particular, e para a humanidade em geral. Sendo assim, o relativo êxito dos programas de documentação investigados nos compeliu a fazer uma proposta básica de projeto de um arquivo linguístico digital para o Setor de Linguística do Museu Nacional/UFRJ com a intenção de que seja, futuramente, materializado num sistema web que contribua para que os professores, os pesquisadores, os estudantes e as comunidades de fala possuam um local adequado para o armazenamento, consulta e compartilhamento, online, de todo o conteúdo dos dados primários de suas pesquisas de campo em linguística documental.

**Palavras-chave:** Programas DOBES, DMA e AILLA. Línguas indígenas ameaçadas. Uma abordagem crítica. Projeto de arquivo linguístico

## ABSTRACT

This dissertation has as a general objective to investigate, critically, the impact of the language documentation programs Dokumentation Bedrohter Sprachen (DOBES), Digital Media Archive (DMA) and Archive of Indigenous Languages of Latin America (AILLA) from their structuring processes to address the problem related to endangered indigenous languages. For this purpose, the idea is to assess, broadly, how the set of initiatives adopted can be validated or not as efficient policies around the desired interests. In terms of theoretical referential, we use material from the most recent publications related to our research object. About the methodology, we adopt the essential elements divided into categories according to: the purpose (applied research), the approach (qualitative- quantitative research), the method (hypothetical-deductive) and the procedures (bibliographical research). About the result of what was analyzed about the object, it can be stated that the linguistic documentation programs studied act with professionalism, are consolidated and have reached their goals in terms of documentation, storage and availability of linguistic archives to the general public. The impact of the DOBES, DMA, and AILLA programs on the issue of endangered indigenous languages is positive, as they can be considered relevant in the effort to avoid an irreparable linguistic and cultural catastrophe for indigenous peoples in particular, and for humanity in general. Thus, the relative success of the documentation programs investigated has compelled us to make a basic project proposal of a digital linguistic archive for the Linguistics Sector of the National Museum/UFRJ, with the intention of having it, in the future, materialized in a web system that will contribute to provide teachers, researchers, students and speech communities with an adequate place for storing, consulting and sharing, online, all the content of the primary data from their field research in documentary linguistics.

**Keywords:** DOBES, DMA and AILLA programs. Endangered indigenous languages. A critical approach. Linguistic archive Project

## LISTA DE QUADROS

<b>Quadro I</b> – Escala EGIDS – análise de perpetuação intergeracional	26
<b>Quadro II</b> - Populações indígenas e número de línguas indígenas na América do Sul	36
<b>Quadro III</b> - População indígena pré-cabralina adaptada de Denevan	42
<b>Quadro IV</b> - População indígena pré-cabralina adaptada de Hemming	42
<b>Quadro V</b> - Padrão básico de documentação linguística	56
<b>Quadro VI</b> - Padrão ampliado de documentação linguística	56
<b>Quadro VII</b> - Funcionalidades, permissões e público-alvo dos módulos arquivista e depositante	142
<b>Quadro VIII</b> - Funcionalidades, permissões e público-alvo do módulo público	142
<b>Quadro IX</b> - Classificação dos eventos	147
<b>Quadro X</b> - Lista dos principais eventos executados	148
<b>Quadro XI</b> - Entidade usuário do sistema	150
<b>Quadro XII</b> - Entidade coleção linguística	151
<b>Quadro XIII</b> - Entidade depósito na coleção	152
<b>Quadro XIV</b> - Entidade família linguística	153
<b>Quadro XV</b> - Entidade língua pesquisada	154

## LISTA DE GRÁFICOS

**Gráfico I** - Percentual de concentração das línguas por continente

27

## LISTA DE FIGURAS

<b>Figura I</b> - Mapa de localização dos projetos DOBES	14
<b>Figura II</b> - Mapa de localização dos projetos ELDP	15
<b>Figura III</b> - As línguas mais faladas na América Latina	38
<b>Figura IV</b> - Fluxo de trabalho na descrição e documentação linguística	57
<b>Figura V</b> - Fluxo de participantes em trabalho de campo linguístico	63
<b>Figura VI</b> - Esquema de corpus IMDI (DOBES) sobre a organização hierárquica de recursos	85
<b>Figura VII</b> - Abrangência do código de conduta DOBES sobre a estrutura organizacional	95
<b>Figura VIII</b> - Níveis de abstração de dados	137
<b>Figura IX</b> - Abstração da coleta de requisitos e da sequência de aplicação dos modelos	138
<b>Figura X</b> - Diagrama de contexto para as operações do arquivista	145
<b>Figura XI</b> - Diagrama de contexto para as operações dos depositantes	145
<b>Figura XII</b> - Diagrama de contexto para as operações dos usuários	146
<b>Figura XIII</b> - Estrutura entidade-relacionamento (ER) no banco de dados	155
<b>Figura XIV</b> - Estrutura hierárquica do módulo arquivista	157
<b>Figura XV</b> - Estrutura hierárquica do módulo depositante	158
<b>Figura XVI</b> - Estrutura hierárquica do módulo usuário	159
<b>Figura XVII</b> - Tela de interface da página principal	161
<b>Figura XVIII</b> - Tela de interface do login de usuário	162
<b>Figura XIX</b> - Tela de interface do cadastro de usuário	163
<b>Figura XX</b> - Tela de interface do cadastro de coleção	164
<b>Figura XXI</b> - Tela de interface do depósito na coleção	165

## LISTA DE SIGLAS

- ABPI** - Associação Brasileira de Propriedade Intelectual
- AILLA** - Archive of Indigenous Languages of Latin America
- ALD** - Arquivo Linguístico Digital
- BD** - Banco de Dados
- CAPES** - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
- CDC** - Código de Conduta
- CEDI** - Centro Ecumênico de Documentação e Informação
- CELP** - Committee on Endangered Languages and their Preservation
- CLAS** - Committee on Latin America Studies
- CNPI** - Conselho Nacional de Proteção aos Índios
- DEL** - Documenting Endangered Languages
- DEL-DDRIG** - Documenting Endangered Languages - Doctoral Dissertation Research Improvement Grants
- DILA** - Laboratorio de Documentación y Investigación en Linguística Y Antropologia
- DMA** - Digital Media Archive
- DOBES** - Dokumentation Bedrohter Sprachen
- EDAAS** - Electronic Data Archive of Aboriginal Studies
- EGIDS** - Expanding Graded Intergenerational Disruption Scale
- ELAR** - Endangered language Archives
- ELDP** - Endangered Language Documentation Program
- ELF** - Endangered Language Found
- EUROBABEL** - European Science Foundation *Best Analytics Program*
- FEL** - Foundation for Endangered Languages
- FUNAI** - Fundação Nacional do Índio
- FUNAI** - Fundação Nacional do Índio
- FUNASA** - Fundação Nacional de Saúde
- GBS** - Gesellschaft für bedrohte Sprachen
- GIDS** - Graded Intergenerational Disruption Scale
- GNU** - General Public Licence
- IBGE** - Instituto Brasileiro de Geografia e Estatística
- IIAP** - Instituto de Investigaciones para la Amazonia peruana
- IMDI** - ISLE Meta Data Initiative

**IMP** - Instituto Max Planck  
**INPI** - Instituto Nacional de Propriedade Industrial  
**ISA** - Instituto Socioambiental  
**ISLE** - International Standard for Language Engineering  
**LLA** - Language Laboratories and Archives  
**LSA** - Linguistic Society of América  
**MEC** - Ministério da Educação e Cultura  
**MN** - Museu Nacional  
**MPEG** - Museu Paraense Emílio Goeldi  
**NEH** - National Endowment for the Humanities  
**NSF** - National Science Foundation  
**OLAC** - Open Language Archives Community  
**OMC** - Organização Mundial do Comércio  
**ONG** - Organização não governamental  
**PDL** - Programa(s) de documentação linguística  
**PHP** - PHP Hypertext Processor  
**PRODOCLIN** - Projeto de Documentação de Línguas Indígenas  
**PROFLIND** - Mestrado Profissional em Linguística e Línguas Indígenas  
**SGBD** - Sistema de Gerenciamento de Banco de Dados  
**SIL** - Summer Institute of Linguistics  
**SLMN** - Setor de Linguística do Museu Nacional  
**SOAS** - School of Oriental and African Studies  
**TIC** - Tecnologia da Informação e Comunicação  
**TLA** - The Language Archive  
**TRIPS** - Trade Related Aspects of Intellectual Property Rights  
**UC** - Universidade de Chicago  
**UFG** - Universidade Federal de Goiás  
**UFRJ** - Universidade Federal do Rio de Janeiro  
**UNESCO** - Organização da Nações Unidas para a Educação, Ciência e Cultura  
**UT** - Universidade do Texas  
**WIPO** - World Intellectual Property organization  
**WWW** - World Wide Web

## SUMÁRIO

<b>1 – INTRODUÇÃO.....</b>	<b>1</b>
1.1 – Tema.....	3
1.2 – Transformação da documentação linguística.....	4
1.3 – Documentação linguística no século XXI.....	6
1.4 – Investimento, crescimento e consolidação.....	16
1.5 – Justificativa.....	17
1.6 – Problema.....	18
1.7 – Objetivos.....	19
1.8 – Referencial teórico.....	19
1.9 – Metodologia.....	21
<b>2 – REVISÃO DA LITERATURA.....</b>	<b>24</b>
2.1 – Línguas indígenas ameaçadas.....	24
2.1.1 – Ameaça em números.....	24
2.1.2 – Fatores de ameaça.....	29
2.1.3 – Maiores riscos.....	30
2.2 – Línguas indígenas na América do Sul.....	34
2.2.1 – Paradigma da diversidade.....	34
2.2.2 – Fatos e números essenciais.....	35
2.3. Línguas indígenas no Brasil.....	39
2.3.1 – Resultado da conquista.....	39
2.3.2 – Características populacionais.....	41
2.3.3 – Línguas ameaçadas e documentação.....	46
2.4 – Propriedades da documentação linguística.....	47
2.4.1 – Discussão teórica.....	47
2.4.2 – Discussão metodológica.....	60
2.4.2.1 – Trabalho de campo.....	60
2.4.2.2 – Ética.....	60
2.4.2.3 – Direitos.....	64
2.4.2.4 – Pesquisa colaborativa.....	70
2.4.3 – Aspectos linguísticos.....	71
2.4.3.1 – Tipologia.....	71
2.4.3.2 – Tipo de dados.....	72
2.4.4 – Prática documental.....	72
2.4.4.1 – Processamento linguístico.....	74
2.4.4.2 – Arquivo digital.....	76
2.4.4.3 – Depositante.....	77
2.4.4.4 – Usuário.....	79
2.4.4.5 – Arquivista.....	80
2.5 – Programas de documentação.....	88
2.5.1 – DOBES.....	89
2.5.1.1 – Antecedentes.....	89
2.5.1.2 – Recursos.....	91
2.5.1.3 – Arquivo linguístico digital.....	93
2.5.1.3.1 – Código de conduta.....	95
2.5.1.4 – Observações críticas.....	97
2.5.2 – DMA.....	101
2.5.2.1 – Antecedentes.....	101
2.5.2.2 – Recursos.....	103

2.5.2.3 – Arquivo linguístico digital.....	107
2.5.2.3.1 – Código de conduta.....	109
2.5.2.4 – Observações críticas.....	110
2.5.3 – AILLA.....	114
2.5.3.1 – Antecedentes.....	114
2.5.3.2 – Recursos.....	116
2.5.3.3 – Arquivo linguístico digital.....	116
2.5.3.3.1 – Código de conduta.....	118
2.5.3.4 – Observações críticas.....	119
<b>3 - ANÁLISE DE RESULTADOS.....</b>	<b>124</b>
3.1 – Uma abordagem crítica.....	124
3.2 – Arquivo linguístico digital do Setor de Linguística/MN/UFRJ.....	133
3.2.1 – Estrutura do projeto.....	139
3.2.2 – Apresentação.....	140
3.2.3 – Identificação de objetivos.....	141
3.2.4 – Diagramas de contexto.....	144
3.2.5 – Eventos de sistema.....	147
3.2.6 – Estrutura de dados.....	148
3.2.7 – Interface sistema-usuário.....	156
3.2.7.1 – Estrutura hierárquica dos módulos.....	157
3.2.7.2 – Modelos de tela na interface sistema-usuário.....	160
3.3 – Considerações finais.....	168
<b>REFERÊNCIAS.....</b>	<b>172</b>
ANEXO A - Kit básico de equipamento de campo.....	192
ANEXO B - Fluxograma etapas da documentação linguística no trabalho de campo	193
ANEXO C - Guia DOBES para gravação de áudio.....	194
ANEXO D - Guia DOBES para gravação de vídeo - parte 1.....	195
ANEXO E - Guia DOBES para gravação de vídeo - parte 2.....	196
ANEXO F - Projetos de documentação DOBES em diversas partes do mundo	197
ANEXO G - Padrões de formato e codificação DOBES – parte 1.....	198
ANEXO H - Padrões de formato e codificação DOBES – parte 2.....	199
ANEXO I - Relação de projetos apoiados pelo DOBES e ELDP no Brasil.....	200
ANEXO J - Línguas indígenas documentadas no MPEG e armazenadas no TLA (DOBES/Max Planck)	201
ANEXO K - Conteúdo parcial das coleções brasileiras no DMA - parte 1.....	202
ANEXO L - Conteúdo parcial de coleções brasileiras no DMA - parte 2.....	203
ANEXO M - Resultado parcial de busca por “Brazil” no AILLA.....	204
ANEXO N - Resultado parcial de busca por “Mbyá Guarani” no AILLA.....	205
ANEXO O – Coleção de língua indígena armazenada no AILLA.....	206

## 1-INTRODUÇÃO

Esta dissertação tem como proposta uma análise crítica e comparativa sobre a estrutura de funcionamento dos seguintes programas de documentação linguística (PDL): *Dokumentation Bedrohter Sprachen* (DOBES), *Digital Media Archive* (DMA) e *Archive of Indigenous Languages of Latin America* (AILLA), vinculando-os às respectivas instituições responsáveis, Instituto Max Planck, Universidade de Chicago e Universidade do Texas. Tal procedimento avalia de que modo os recursos financeiros, acadêmicos, documentais e tecnológicos empregados, por essas organizações, impactam a situação de risco enfrentada pelas línguas indígenas ameaçadas. Assim, este texto está dividido em 3 capítulos acompanhados das considerações finais e das referências bibliográficas.

O capítulo 1 apresenta o tema como resultado direto de sucessivos processos de transformação experimentados pela documentação linguística, fenômeno que levou, a partir dos anos 1990, ao surgimento de novas teorias e metodologias na área, o que permitiu outras possibilidades de se fazer documentação. Tal iniciativa, protagonizada por acadêmicos preocupados com a ameaça à diversidade linguística, rapidamente vinculou-se às grandes instituições de pesquisa, financiamento e universidades que ensejaram o surgimento de programas de documentação associados a seus sustentáculos informáticos denominados arquivos linguísticos.

O capítulo 2 está dividido em cinco seções, as quais são detalhadas abaixo:

- A primeira seção apresenta um quadro geral da grave ameaça à diversidade linguística, ilustrando-o com números relativos aos fatores que representam os maiores riscos, com dados fornecidos por instituições internacionais envolvidas com as discussões levantadas.
- A segunda seção aborda, nesse contexto de preocupação, os primeiros estudos sobre a situação das línguas indígenas na América do Sul na totalidade.
- A terceira seção investiga a situação das línguas indígenas no Brasil, abordando diversas questões como população, distribuição geográfica, famílias maiores e menores e a importância da documentação linguística nesse cenário de línguas ameaçadas.
- A quarta seção discute questões referentes a aspectos teóricos, metodológicos e práticos da documentação linguística. Sobre este

último ponto, são debatidos temas como: trabalho de campo (ética, direitos, pesquisa colaborativa), aspectos linguísticos (tipologia, tipo de dados) e prática documental (criação de arquivo digital, papéis e metas de depositantes, usuários e arquivistas).

- A quinta seção reserva a análise crítico-comparativa dos três programas de documentação que são objeto desta dissertação: DOBES (Instituto Max Planck), DMA (Universidade de Chicago) e AILLA (Universidade do Texas). Nessa abordagem, fazemos uma análise sobre as estruturas gerais que os sustentam, expondo alguns de seus projetos de documentação na América Latina, dedicando, nesse sentido, especial atenção aos seus arquivos linguísticos digitais.

O capítulo 3 prossegue com os apontamentos e reflexões críticas. Nesse sentido, a análise de resultado é um instrumento importante ao relacionar as descobertas resultantes de nossos estudos com a resposta ao problema da pesquisa. Mas não se resume a isso, os conhecimentos adquiridos e as descobertas sobre o tema são amplamente avaliados, principalmente, aquelas que suscitaram dificuldades, dúvidas, novidades ou omissões em relação aos questionamentos que formularam o problema. Tais discussões, com destaque para os principais aspectos estruturantes dos programas DOBES, DMA e AILLA, nos permitiram elaborar uma proposta básica de projeto de arquivo linguístico digital para o Setor de Linguística do Museu Nacional (SLMN). Essa iniciativa está vinculada aos critérios de exigência da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) para os cursos de mestrado na modalidade profissional. Nas considerações finais, fazemos um balanço geral dos resultados alcançados pela pesquisa, em torno do problema, dos objetivos e da metodologia aplicada. Abordamos também as lacunas que persistem e podem ser respondidas em trabalhos futuros, assim como as principais dificuldades enfrentadas ao longo do desenvolvimento desta atividade acadêmica, entre outras questões.

Feito esse breve preâmbulo, passemos às nossas investigações.

## 1.1 – Tema

Esta dissertação visa desenvolver, com base nos referenciais teóricos, a análise crítica dos métodos e práticas utilizadas pelos programas de documentação linguística de modo a avaliar qual o resultado de suas ações sobre a difícil realidade de ameaça à diversidade linguística. Como a quantidade de projetos é imensa e engloba dezenas de programas responsáveis por arquivos linguísticos digitais, optamos por três (DOBES, DMA e AILLA), destacando os projetos desenvolvidos na América do Sul.

O crescimento acelerado da documentação linguística<sup>1</sup>, como subárea da Linguística, nos últimos vinte anos, não deve ser dissociado da expansão, em paralelo, dos arquivos digitais e as respostas para isso podem ser vistas ao longo desta pesquisa. Esperamos que a entrega deste estudo contribua para que o conhecimento da teoria, metodologia, material tecnológico e o uso da ética no campo estimulem pesquisadores ainda reticentes na adoção desses poderosos recursos. Algo que se justifica até aqui pelas dificuldades impostas pelas exíguas produções acadêmicas voltadas para os leigos ou iniciantes interessados em se aventurar no tema. Entretanto, é possível afirmar, a partir do esforço implícito na criação deste trabalho, essencialmente bibliográfico, que produções correlatas existem e são devidamente referenciadas.

Inicialmente, a metodologia desta dissertação propunha pesquisa de campo associada a estudo de caso visando implementar elementos teóricos, metodológicos e técnicos da documentação linguística *in loco*, com vistas a avaliar, após a coleta de dados, o resultado de sua aplicação nos eventos da vida cotidiana da comunidade de fala. Tal impossibilidade impactou em todo projeto de pesquisa original e na sua consecução na totalidade. Contudo, as explicações são plenamente compreensíveis frente às graves consequências da crise sanitária de COVID-19 que abateu o mundo por mais de dois anos, contabilizando milhões de mortes, com efeitos deletérios sobre diferentes grupos étnicos, inclusive indígenas, que vivem o resultado desse flagelo acompanhado de outras crises cotidianas.

Na tentativa de nos solidarizarmos com todas as populações indígenas do

---

<sup>1</sup> A partir daqui, até o final desta dissertação, sempre que houver referência ao termo “documentação linguística”, no âmbito dos programas de documentação analisados, deve-se inferir que é no formato digital.

mundo e compensar as perdas assumidas por esta produção acadêmica, apresentamos os presentes estudos com o maior aprofundamento teórico dos pontos que contribuirão para iluminar os pesquisadores interessados na utilização dos recursos da documentação linguística. Essa estratégia soma-se a tecnologia dos arquivos linguísticos ligados aos programas de documentação para enfrentar o apagamento da cultura imaterial dos povos originários.

## 1.2 - Transformação da documentação linguística

O conceito e o modo de documentar vêm mudando ao longo do tempo e por isso consideramos importante tratar, brevemente, como esse procedimento vem sendo realizado e os fatores que determinaram as transformações profundas que linguistas, antropólogos e outros especialistas presenciaram desde o final do século XIX até a atualidade. Esse período pode ser dividido em quatro etapas, conforme os recortes temporais assinalados abaixo (HENKE & BEREZ-KROEKER, 2016; DA SILVA, 2018).

Em termos de documentação, registrada em materiais tradicionais como o papel, os cem anos que separam o final do século XIX e a década de 1990 se caracterizam pelo pioneirismo da documentação das línguas indígenas americanas efetuada por estruturalistas como Franz Boas e Edward Sapir. Os materiais produzidos (notas de campo, traduções, dados de elicitación, compilações lexicais e descrições gramaticais) foram armazenados em locais diversos, desde arquivos<sup>2</sup> particulares até repositórios de museus e universidades onde o acesso era restrito a pesquisadores dispostos a viagens para a análise dos materiais (GOLLA, 1995; JOHNSON, 2004 apud HENKE & BEREZ-KROEKER, 2016).

Até a segunda metade do século XX prevaleceu a utilização de diversas instituições para armazenar o trabalho de documentação dos pesquisadores, entre as quais destacam-se: a *American Philosophical Society* (APS) que desde 1945 passou a armazenar a coleção de Boas, resultante da documentação das línguas indígenas norte-americanas; a Universidade da Califórnia (Berkeley), que guarda os trabalhos

---

<sup>2</sup> Conforme Houaiss (2010, p. 67), tomamos o sentido, entre outros, de “arquivo”, como “[...] local ou móvel onde se guardam os documentos [...]”; sendo esse o local tradicional de armazenamento até o surgimento dos arquivos digitais. Em relação aos nossos estudos, o contexto histórico é outro. Trata-se do período em que já é possível criar e armazenar dados linguísticos digitais (p.ex., áudio, vídeo, anotação de gravações, léxicos, notas de campo, etc.) em banco de dados online.

linguísticos de A. L. Kroeber, Pliny Earle Goddard, T.T. Waterman, Edward Sapir e E. W. Gifford; além do grande acervo de pesquisas de línguas indígenas elaborado por pós-graduandos, discípulos boasianos, responsáveis pela produção de gramáticas, dicionários de texto, materiais de áudio e notas de campo arquivados no *Survey of Californian Indian Languages* (WOODBURY, 2011).

A partir da década de 1960 surgem outros arquivos, como o *National Anthropological Archives* (NAA), criado em 1965, onde são guardadas as pesquisas de línguas indígenas norte-americanas, entre as quais destaca-se as feitas por J. P. Harrington de 1915 a 1954 (GOLLA, 1995; MACRI & SARMENTO, 2010 apud HENKE & BEREZ-KROEKER, 2016). Na década de 1970 é fundado o *Alaska Native Language Center* (ANLC), por Michael Krauss, atualmente conhecido como *Alaska Native Language Center* (ANLA), na Universidade do Alasca, que armazena a documentação de materiais impressos e de áudio de 20 línguas indígenas do Alasca (WOODBURY, 2010; HOLTON, 2012 apud HENKE & BEREZ-KROEKER, 2016).

Paulatinamente, o desenvolvimento tecnológico, desde o início do século XX, alterou o trabalho de campo linguístico na forma de armazenagem dos dados. Nesse período tornou-se possível gravar áudio de diversas formas, entre as quais, através de cilindros de cera e fonógrafo (anos 1920-1930), gravação de fita (1950), gravação de vídeo (1980). Esses métodos analógicos de documentação e armazenamento começaram a ser substituídos pela tecnologia digital na segunda metade do mesmo século com o uso de computadores (GOLLA, 1995; THIEBERGER & MUSGRAVE, 2007 apud HENKE & BEREZ-KROEKER, 2016). Os primeiros repositórios voltados às comunidades linguísticas criados dentro dessa nova perspectiva foram o *Oxford Text Archive* (1976) e o *Linguistic Data Consortium*, elaborado pela Universidade da Pensilvânia (1992).

Na década de 1990, o avanço tecnológico digital e a queda nos custos dos equipamentos de informática coincidiu com dois fenômenos importantes: o surgimento de novos projetos de documentação a partir das abordagens propostas pela linguística documental e a preocupação dos linguistas em relação à ameaça de desaparecimento de diversas línguas minoritárias. Diante disso, ocorreram várias reações, entre as quais a constatação de que “[...] obviamente devemos repensar seriamente nossas prioridades para que a Linguística não entre na história como a única ciência que presidiu inconscientemente o desaparecimento de 90% do próprio campo ao qual se dedica [...]” (KRAUSS, 1992, p. 10). Diante dessa preocupação, Himmelmann (1998)

trouxe sua proposta de documentação linguística como subárea separada da linguística estrutural<sup>3</sup>. Isso para alguns nada mais representou do que o retorno aos primórdios da disciplina no que se refere à primazia da pesquisa de campo, tendo em vista que “a linguística convencional se tornou cada vez mais teórica desde a revolução generativa das décadas de 1950 e 1960” (HIMMELMANN, 2006; CONATHAN, 2011; WOODBURY, 2003 apud HENKE & BEREZ-KROEKER, 2016). Apesar dessas controvérsias, o interesse e as ações voltadas à documentação da diversidade linguística têm crescido apoiadas nas propostas da documentação linguística.

### 1.3 - Documentação linguística no século XXI

O início do século XXI marca um conjunto de transformações no processo de documentação linguística diretamente relacionadas ao desenvolvimento de novos modelos, práticas, apoio financeiro a projetos e ao avanço descomunal da tecnologia da informação (TI) que mudam o fluxo de trabalho e permitem seu crescimento e consolidação.

Primeiramente, como mencionado, o fator principal é a ameaça às línguas minoritárias. A diminuição da diversidade linguística e as consequências desse fenômeno para falantes e linguistas têm sido motivo de debates e reações, supra e infranacional, no meio acadêmico e na sociedade civil. Como afirma Mufwene (2004, p. 3): “[...] as publicações linguísticas sobre a ameaça e morte das línguas aumentaram desde a publicação de Dorian (1989), *Investigating Obsolescence*, e mais ainda desde a publicação, em 1992, de uma edição do especial da *Language* (The Linguistic Society of America’s Journal) sobre o assunto (*Language*, vol. 68, no. 1, 1992) [...].”

---

<sup>3</sup> Os conceitos de estruturalismo linguístico ou linguística estrutural fazem parte da teoria linguística que interpreta a linguagem como um sistema independente, com regras próprias, cujos elementos linguísticos definem-se conforme o tipo de relacionamento que possuem entre si. Desenvolvida pelos seguidores de Saussure, após sua morte (1857-1913), a linguística estrutural baseia-se na teoria de análise linguística que pressupõe a “explicação da língua por si mesma”. Com esse objetivo, busca-se a verificação das relações que unem os componentes no discurso com vistas a determinar a funcionalidade dos diferentes tipos de relações estabelecidas. Assim, considera-se a língua uma estrutura formada por um conjunto de elementos onde cada um possui distinto valor funcional (PETTER, p. 14).

A partir de uma interpretação distinta, Austin & Sallabank (2011, p. 1), identificam, no fim da Segunda Guerra Mundial, uma profunda diminuição de línguas faladas por autóctones em regiões como a Sibéria, a Ásia e as Américas. Nesse cenário, segundo eles, são os mais velhos que mantêm a fala da língua ameaçada, enquanto os mais jovens passam a utilizar a língua majoritária regional, nacional ou global. A substituição da língua pode acontecer rapidamente, até duas gerações, ou gradual e continuamente por gerações. Esse fenômeno de mudança da língua se dá nos períodos de bilinguismo ou multilinguismo instáveis, em que os falantes usam duas ou mais línguas, porém uma delas é mais dominante e usada de modo cada vez mais intensivo até que assume o lugar que antes pertencia à língua ameaçada de extinção. Esse tipo de evento, que aconteceu em várias frentes, vem alarmando os linguistas desde a década de 1990.

O volume de trabalhos acadêmicos lançados no bojo dessa temática e preocupações é vasto e significativo, entre os quais destacam-se as obras de pesquisadores como Fishman (1988, 1991), Adelaar (1991), Robins & Uhlenbeck (1991), Brenzinger (1992, 1998, 2015), Hagège (1992, 2000), Hale (1992), Mühlhäusler (1996), Cantoni (1997), Dixon (1997), Grenoble & Whaley (1998, 2006), Grinevald (1998), Hazaël-Massieux (1999), Reyhner et al. (1999), Crystal (2000), Nettle & Romaine (2000), Skutnabb-Kangas (2000), Hinton & Hale (2001), Maffi (2001), Mufwene (2001), Swaan (2001), Dalby (2002), Harmon (2002), Joseph et al. (2003), Maurais & Morris (2003), Phillipson (2003), Harrison (2008), Austin & Sallabank (2011), Crevels (2012), Sallabank (2013), entre outras publicações.

A produção acadêmica sobre o assunto não termina com as referências citadas, mas permitem mostrar sua relevância. Como visto, há décadas, linguistas do mundo inteiro refletem e buscam soluções para a questão das línguas ameaçadas a partir de uma abordagem político-acadêmica. Eventos de envergadura, ao nível internacional, têm ocorrido, como o Simpósio de Línguas Ameaçadas, denominado *Linguistic Society of America*<sup>4</sup>, assim como a declaração feita no *XVe Congrès international des linguistes*<sup>5</sup>. Dos encontros de 1991-1992, Hale et al. (1992) fez um

---

<sup>4</sup> Evento realizado em Chicago, Illinois, entre 3 e 6 de janeiro de 1991. Confira mais informações no relatório do encontro disponível em: <Ling. Society of America 1991.pdf (linguisticsociety.org)>. Acesso em: 16 jun. 2021.

<sup>5</sup> Evento realizado em Québec, Canadá, entre 9 e 14 de agosto de 1992. Confira mais informações nas atas do encontro disponíveis em: <Actes du XVe Congrès international des linguistes: Québec, Université Laval, 14/09/1992: les langues menacées = Processo do XVº Congresso Internacional de

alerta para o alto grau de ameaça que as línguas minoritárias sofrem, em denúncia em que conclama o empenho dos linguistas para o registro das línguas mais ameaçadas e a criação de arquivos linguísticos como referências futuras. Outros linguistas como Krauss (1992), Nettle & Romaine (2000) e Crystal (2000) referendaram as mesmas preocupações com estimativas sombrias para o futuro dessas línguas se medidas mais contundentes não fossem tomadas.

Apesar de projeções, como as fornecidas pelo *Ethnologue*<sup>6</sup> (veja maior detalhamento sobre essas projeções no capítulo 2), indicarem que entre 6000 e 7000 línguas sejam faladas no mundo, as previsões são desanimadoras, ao indicarem que a metade desse número deve desaparecer até o final do século XXI. Austin & Sallabank (2011) estimaram que desse total, a metade deixaria de existir daqui a poucas gerações porque crianças não aprendem mais a língua falada pelos pais como a primeira língua. O fenômeno, para eles, poderia ser explicado pela divisão desigual entre as línguas e a população mundial. Existe um pequeno número de línguas utilizadas pela maioria da população global (inglês, chinês, hindi, espanhol, francês, etc.)<sup>7</sup> com mais de 50 milhões de falantes cada uma e há, por outro lado, inúmeras línguas com pequenas comunidades de falantes (cujo número oscila entre milhares, centenas e dezenas). Nesse caso, seriam determinantes as pressões de ordem política, econômica e social das nações de línguas majoritárias nas situações em que os falantes de línguas minoritárias abandonam a língua de seus ancestrais.

Nesse quadro, a complexa situação das línguas na China e na Índia merecem algumas considerações particulares em relação ao conjunto das 20 línguas mais faladas no mundo. Na China, a língua oficial, o chinês padrão ou mandarim padrão (Putonghua), baseia-se na variedade linguística específica do mandarim falado em Pequim, e possui condição de língua oral oficial da República Popular da China e da

---

Linguistas: Québec, Université Laval, 9-14 agosto de 1992: línguas ameaçadas (Livro, 1993) [WorldCat.org]>. Acesso em: 16 jun. 2021.

<sup>6</sup> O *Ethnologue* (Languages of the World) é resultado da publicação de referência anual (impressa e online) sobre estatísticas globais e outras informações das línguas vivas. A primeira publicação foi feita em 1951 até tornar-se uma publicação anual feita pela SIL International, organização cristã estadunidense. Segundo os editores, os principais objetivos são o estudo, o desenvolvimento e a documentação das línguas para promover a tradução da Bíblia em diferentes idiomas e a alfabetização religiosa.

<sup>7</sup> O inglês é a língua mais utilizada no mundo se forem considerados os falantes nativos e não nativos. Por outro lado, se forem contabilizados somente os falantes nativos, o chinês Mandarim toma essa posição. Disponível em: <<https://www.ethnologue.com/guides/most-spoken-languages>>. Acesso em: 15 ago. 2022.

República da China; além de ser uma das línguas oficiais de Singapura e uma das seis línguas oficiais das Nações Unidas. O país, com população de 1,4 bilhão de pessoas, fala as línguas que fazem parte da grande família linguística sino-tibetana (70% da população chinesa fala mandarim e outros 30% falam variedades linguísticas). Conforme THOMASON (2001) e CHAPPEL (2016), esta família se divide em dois ramos linguísticos: 1) línguas siníticas (termo linguístico que se refere às línguas e variedades linguísticas chinesas que, em alguns casos, são muito diferentes entre si, ou seja, ininteligíveis), e 2) línguas tibeto-birmanesas. As línguas e variedades linguísticas que compõem estes ramos linguísticos fazem parte, portanto, do cotidiano comunicacional da área que se expande deste o leste até o sudeste asiático. Nesse cenário, enquanto as línguas siníticas (línguas chinesas) são faladas por 91% da população, as línguas tibeto-birmanesas representam o ramo que possui o maior número de línguas (grande parte delas faladas fora da China), mas não o maior número de falantes se comparada às línguas pertencentes à família sino-tibetana (CHAPPEL, 2016; THOMASON, 2001). O percentual populacional restante, em torno de 8%, é composto pelas 56 minorias étnicas, cada uma delas com populações proporcionalmente semelhantes a pequenos países europeus, algumas das quais, a título de exemplo, com números significativos de falantes de variedades linguísticas, como: Wu (77 milhões), Yue (71 milhões), Min (60 milhões)<sup>8</sup>, entre outras (LEWIS et al., 2014).

Em relação à Índia, constitucionalmente, a língua oficial é o hindi. Porém, o inglês (o artigo 343 deu ao Parlamento a possibilidade de manter o uso continuado da língua para fins oficiais), língua predominante do período colonial, ainda se impõe devido a sua situação de maior língua franca internacional do mundo. Por demanda de seu aprendizado é exigido por alguns setores da sociedade, como aqueles ligados ao comércio internacional e à área científica. O país, de forte multilinguismo, possui uma população de 1,2 bilhão de pessoas<sup>9</sup>. De acordo com Thomason (2001), a maioria das línguas indianas fazem parte de duas grandes famílias linguísticas, a indo-europeia e a dravidiana. Em termos percentuais, enquanto o indo-europeu é falado por 73% da população, o dravidiano é falado por 24% e os 3% restantes são falantes

---

<sup>8</sup> Cf. University of Leeds. Disponível em: <[https://commons.wikimedia.org/wiki/File:Map\\_of\\_sinitic\\_languages\\_full-en.svg](https://commons.wikimedia.org/wiki/File:Map_of_sinitic_languages_full-en.svg)>. Acesso em: 20 set. 2022.

<sup>9</sup> Cf. National Portal of India. Disponível em: <<https://www.india.gov.in/india-glance/profile>>. Acesso em: 21 set. 2022.

de línguas pertencentes a diversas famílias menores, como a austro-asiática, a tibeto-birmanesa, entre outras. A família indo-europeia reúne vários sub-ramos, entre os quais o indo-ariano, que abriga as línguas indianas (SENGUPTA & SAHA, 2016), como o hindi, o urdu (língua oficial do Paquistão), marata, bengali, o shina, o assamês e o romani. Essas línguas são faladas principalmente na Índia, mas também são largamente utilizadas por imigrantes em outras partes da Ásia, na África, no Pacífico, no Caribe, e outros locais (THOMASON, 2001).

Devido à importância das línguas nativas minoritárias como um elemento identitário e cultural de um povo, seu desaparecimento pode provocar perdas irreparáveis para a humanidade, e essa possibilidade tem preocupado várias instituições internacionais. Uma delas é a Organização das Nações Unidas para a Educação, Ciência e Cultura (UNESCO)<sup>10</sup> ao destacar que a cultura, a língua e demais características de um povo fazem parte de um conjunto de valores que compõem um “modo de vida” e extrapolam a ideia de “acumulação de obras e conhecimentos produzidos”. Assim, esses elementos impõem-se por meio de outras linguagens como a música, a dança, a mitologia, os ritos, o artesanato, em circunstâncias em que os saberes e a cosmovisão não são transmitidos exclusivamente pela oralidade. Ao corroborar essas afirmações, ressaltamos que esses aspectos da cultura tradicional fazem parte do patrimônio universal da humanidade e contribuem para a afirmação da identidade cultural das minorias, inclusive as indígenas. O legado cultural imaterial resultante dessa diversidade é reproduzido, constantemente, em elementos socioculturais perceptíveis nesses grupos através de sua história, na expressão de seus corpos, na relação estabelecida entre língua-linguagem, além de serem elementos determinantes de sua territorialidade. Assim, podemos afirmar que os direitos linguísticos fazem parte do arcabouço de direitos humanos fundamentais. Conforme Chiriboga (2006), o outro setor preocupado com a questão em tela é a comunidade acadêmica ligada aos instrumentos jurídicos internacionais que preveem a garantia de direito à identidade cultural aos povos indígenas e às minorias nacionais conforme o sistema interamericano. Nesse caso, o papel do Estado é crucial como garantidor desses direitos nos territórios onde os grupos étnicos minoritários estiverem situados.

Sobre a questão das línguas minoritárias, Grenoble & Singerman (2017, p. 8)

---

<sup>10</sup> Cf. Prefácio da Declaração Universal da Unesco sobre a diversidade cultural (2001).

afirmam que o “termo língua minoritária” possui uma definição básica, em termos numéricos e geográficos (falada por pelo menos 50% da população em determinada região, estado ou país). Por outro lado, possui também definições fluidas se forem consideradas as populações de imigrantes cuja língua é majoritária no país de origem, mas torna-se minoritária no país de destino. A título ilustrativo, os autores citam o caso do espanhol, língua majoritária em diversos países, principalmente latino-americanos, porém minoritária, de modo geral, nos Estados Unidos.

Mesmo em território estadunidense a língua espanhola sofre diferentes gradações no âmbito das línguas minoritárias, pois consoante o território ocupado pelas populações latinas, a posição valorativa dessa língua é alterada. Ao ressaltar a relevância das reflexões tratadas acima, os autores abordam a ocultação de fatores explicativos fundamentais para uma compreensão mais precisa da questão relacionada a essas línguas. Para isso, eles chamam a atenção para três pontos cruciais: “o prestígio econômico, social ou político de uma língua minoritária”.

Esses elementos, na interpretação dos autores, podem contribuir para enriquecer o quadro explicativo quando se trata de analisar o tema em questão, ao mencionarem duas situações distintas que comportam essa classificação: o caso de países denominados “potências” que observam o abandono dessa posição hegemônica a partir do momento em que seus cidadãos transformam-se em imigrantes; e a situação das línguas nativas, que se tornam permanentemente constrangidas, em termos de prestígio, pelas línguas oficiais.

Os debates acima ocorreram em meio ao crescente fluxo financeiro destinado a investimentos em projetos de linguística documental conduzidos por instituições seriamente preocupadas com as línguas ameaçadas (quadro que engloba 90% das 7.000 línguas faladas). A mesma ação foi estendida para o treinamento de falantes temerosos com a possibilidade de desaparecimento de suas línguas maternas (AUSTIN, 2007; WHALEN, 2004). Tais iniciativas surgiram no bojo das percepções sobre a gravidade do problema, debatidas no final da década de 1980<sup>11</sup>, que coincidiram com a paulatina facilitação do acesso às mídias, tecnologias de informação e ao arquivamento como ferramentas fundamentais à coleção, análise,

---

<sup>11</sup> Conforme Himmelmann (2008, p. 339) apud Austin (2015): “o gatilho foi uma breve apresentação de Johannes Bechert no 14º Congresso Internacional de Linguistas em Berlim Oriental ocorrido em 1987 e uma moção elaborada por Christian Lehmann apresentada à reunião de negócios do Comité International Permanent des Linguistes (CIPL), instando a comissão a tomar medidas para trazer a questão das línguas ameaçadas à atenção dos linguistas profissionais e do público.”

preservação e disseminação dos registros documentais. Essas ações conjugam-se, em termos éticos, com a garantia dos direitos fundamentais das comunidades de fala, participantes no processo de documentação das próprias línguas colaborativamente. Nesse quadro, os recursos liberados para os projetos têm dois impactos: primeiro, contribuem para a consolidação da linguística documental e, segundo, permitem a elaboração de uma diversidade de projetos conforme os parâmetros determinados pelos financiadores.

O acesso a recursos voltados à documentação linguística levou alguns pesquisadores a enquadrarem-se em determinada perspectiva documental. Assim, da segunda metade dos anos 1990 até 2010, registra-se aporte financeiro significativo feito por agências governamentais e não governamentais, entre as quais destacam-se: o Programa DOBES<sup>12</sup>, financiado pela *Volkswagen Foundation*, que no período citado investiu milhões de euros em dezenas de projetos de pesquisa (ver Figura I); o *Endangered Language Documentation Program (ELDP)*<sup>13</sup>, ligado a *Arcadia Fund*, que liberou 7,25 milhões de libras e financiou 195 projetos e 350 bolsas (ver Figura II); o *Documenting Endangered Languages (DEL)*<sup>14</sup>, vinculado à *National Endowment for the Humanities (NEH)* e à *National Science Foundation (NSF)*, agências do governo norte-americano que investiu 10 milhões de dólares em 320 projetos e 60 bolsas; e o projeto da *European Science Foundation Best Analytics Program (EuroBABEL)*, que liberou 8 milhões de euros. Outros recursos devem ser considerados, como os oriundos do *Endangered Language Found (ELF)*<sup>15</sup>, da *Foundation for Endangered*

---

<sup>12</sup> O Programa DOBES (“Dokumentation bedrohter Sprachen”) tem início em 2000, a partir de iniciativa da Fundação Volkswagen em documentar línguas ameaçadas de extinção no curto prazo. Sua política destinada à arrecadação de fundos foi encerrada em 2011. Disponível em: <<https://dobes.mpi.nl/dobesprogramme/?lang=pt>>. Acesso em: 25 jul 2021.

<sup>13</sup> O ELDP foi fundado em 2002 pelo Fundo Arcádia (Arcadia Fund) e tem ligação com fundo de caridade de Lisbet Rausing e Peter Baldwin. Intitulam-se guardiões do “patrimônio cultural e dos ecossistemas ameaçados”. Informam que o acesso a sua base de dados é aberto, sem barreiras e custos aos interessados. Disponível em: <<https://www.eldp.net/en/about+us/>>. Acesso em: 25 jul. 2021.

<sup>14</sup> O programa Documenting Endangered Languages (DEL) é uma parceria entre o National Endowment for the Humanities (NEH) e a National Science Foundation (NSF), as duas agências federais foram fundadas em 1965 no governo Lyndon Johnson para desenvolver e avançar o conhecimento sobre línguas ameaçadas. Disponível em: <<https://www.neh.gov/grants/preservation/documenting-endangered-languages>>. Acesso em: 26 jul. 2021.

<sup>15</sup> Criada em 1996, a ELF apoia projetos de preservação e documentação de línguas ameaçadas de extinção. O apoio ao trabalho sobre línguas ameaçadas ocorre na forma de subsídios para indivíduos, tribos e museus. As bolsas da ELF já promoveram esse trabalho em mais de 60 países. Disponível em: <<http://www.endangeredlanguagefund.org/about-elf.html>>. Acesso em: 26 jul. 2021.

*Languages* (FEL)<sup>16</sup>; do Gesellschaft für bedrohte Sprachen (GBS)<sup>17</sup>; e da Unesco<sup>18</sup>, importantes para diversos projetos de formação e criação de métodos de pesquisa (AUSTIN, 2015, 2010; WOODBURY, 2003).

Os resultados desses investimentos refletem o incremento, entre os anos de 2009 e 2010, no número de alunos com mestrado e doutorado na área em universidades europeias e norte-americanas. Este crescimento proporcionou o aumento de pesquisadores (seniores e bolsistas de pós-doutorado) que, mediante workshops, conferências e sucessivas publicações, têm contribuído para a consolidação dos programas de documentação e da documentação linguística como uma importante subárea da Linguística (AUSTIN, 2015). Dentro de nossos estudos analisaremos, detalhadamente, os programas DOBES, DMA e AILLA e seus respectivos acervos linguísticos digitais. Para isso, reservamos os capítulos 2 e 3. Abaixo, seguem os mapas globais de projetos dos programas DOBES e ELDP.

---

<sup>16</sup> A Foundation for Endangered Languages (FEL) é uma organização sem fins lucrativos, registrada como Charity na Inglaterra e no País de Gales, fundada em 1996. Seu atual presidente é Nicholas Ostler. Existe para oferecer suporte, habilitar e auxiliar a documentação, proteção e promoção de idiomas ameaçados. Disponível em: <<http://www.ogmios.org/manifesto/index.php>>. Acesso em: 28 jul. 2021.

<sup>17</sup> A Gesellschaft für bedrohte Sprachen (Sociedade para Línguas Ameaçadas), criada em 1999, é sediada em Colônia, na Alemanha e visa organizar e apoiar o uso, manutenção e documentação de línguas e variações linguísticas ameaçadas. Disponível em: <<http://gbs.uni-koeln.de/wordpress/index.php/en/welcome/>>. Acesso em: 28 jul. 2021.

<sup>18</sup> Cf. Levantamento da UNESCO sobre as línguas ameaçadas no “UNESCO Atlas of the World's Languages in Danger.” Disponível em: <http://www.unesco.org/languages-atlas/>. Acesso em: 28 jul. 2021.



Figura I: Mapa de localização dos projetos DOBES<sup>19</sup>

<sup>19</sup> Figura retirada e adaptada de DOBES. Disponível em: <https://dobes.mpi.nl/?lang=pt>. Acesso em: 19 mar. 2023.



Figura II: Mapa de localização dos projetos ELDP<sup>20</sup>

<sup>20</sup> Figura retirada e adaptada de ELDP. Disponível em: <https://www.eldp.net/en/our+projects/map/>. Acesso em: 19 mar. 2023.

#### 1.4 – Investimento, crescimento e consolidação

O desenvolvimento da moderna documentação linguística tem relação direta com o avanço acelerado da Tecnologia da Informação e Comunicação (TIC)<sup>21</sup>. Essa mudança se reflete na criação, em 1991, no *Australian Institute of Aboriginal and Torres Strait Islander Studies* (AIATSIS)<sup>22</sup>, daquele que é considerado o primeiro arquivo digital a tratar de línguas ameaçadas no mundo, o *Electronic Data Archive of Aboriginal Studies* (EDAAS) (THIEBERGER, 1994 apud HENKE & BEREZ-KROEKER, 2016, p. 416)<sup>23</sup>. A ligação entre a TIC e a linguística documental tem reflexo direto em suas práticas, algo que veremos detalhadamente adiante, mas podemos indicar desde já a possibilidade de gravação mais eficiente e maior qualidade de áudio e vídeo; maior velocidade de processamento, análise e armazenamento dos materiais; ampliação da distribuição e divulgação dos materiais produzidos no ambiente virtual, principalmente na Internet, utilizando formato<sup>24</sup> e extensão de arquivo que facilitam a acessibilidade das informações.

Como vimos, os atuais princípios da documentação linguística moderna têm sua origem em passado recente que nos remete às práticas de Boas e Sapir na passagem do século XIX para o século XX, no que tange a registro e armazenamento de material físico. A década de 1990, por outro lado, indica o impacto do avanço tecnológico com o acesso facilitado aos equipamentos informáticos e o foco nas línguas ameaçadas e em perigo de extinção. Deste modo, a junção desses eventos permitiu e levou a grandes transformações no universo da Linguística.

Em termos tecnológicos, tornou-se possível catalogar, registrar e preservar dados primários de forma sistematizada e científica por meio de recurso próprio e objetivos definidos. Em reação aos riscos sobre a diversidade linguística, pode-se afirmar que as comunidades de fala têm se tornado protagonistas nos projetos de documentação, mediante treinamento para a criação de dados de corpora e utilização dos equipamentos de mídia. Elas têm produzido sua própria documentação, cujos

<sup>21</sup> O termo se refere ao conjunto de recursos tecnológicos reunidos para o processamento de informações tornando possível o envio e o recebimento de dados digitais (ver SEMENOV 2005).

<sup>22</sup> Cf. AIATSIS. Disponível em: <http://aiatsis.gov.au/>. Acesso em: 19 ago. 2021.

<sup>23</sup> Cf. THIEBERGER, N. Report on the AIATSIS visiting research fellowship, Aboriginal Studies Electronic Data Archive: A report to AIATSIS Council on the conclusion of the Visiting Research Fellowship, 1994.

<sup>24</sup> Conforme Houaiss (2010, p. 369), tomamos o sentido de “formato”, no âmbito da ciência da informação como “[...] padrão magnético criado por quem faz uma formatação (p.ex., de um arquivo, impressora, etc.).”

resultados são a produção de gravações de áudio e vídeo, gramáticas, dicionários, literaturas, etc.

Em resumo, atualmente, encontramos cada vez mais projetos que privilegiam os modelos participativos, onde o aspecto humano é fundamental no que se refere à expansão da interação de pesquisadores, consultores e comunidades de fala no processo de documentação. Utiliza-se, para isso, intensivamente, os recursos informáticos (aspecto científico) que permitem a diminuição das distâncias entre os usuários e os arquivistas responsáveis pelo armazenamento do corpus documental no arquivo digital (DA SILVA, 2018; HENKE & BEREZ-KROEKER, 2016).

Ao vislumbrar uma futura seção de documentação linguística vinculada ao Setor de Linguística do Museu Nacional, esse aspecto participativo da moderna documentação seria uma das características em destaque.

### 1.5 - Justificativa

A escolha do tema “O impacto dos programas DOBES, DMA e AILLA sobre a questão das línguas indígenas ameaçadas: uma abordagem crítica e proposta básica de um projeto de arquivo linguístico para o Setor de Linguística do Museu Nacional” deve-se, principalmente, a dois fatores: o primeiro, está diretamente relacionado, obviamente, à questão da ameaça de extinção de várias línguas indígenas, e o segundo, às consequências negativas desse fenômeno para a diversidade linguística.

O autor desta dissertação envolveu-se com o tema ao perceber que poderia associar a formação acadêmica em ciências humanas com a experiência na área tecnológica. A soma desses fatores foi determinante para o ingresso no Mestrado Profissional em Linguística e Línguas Indígenas do Museu Nacional, em 2018. Além disso, existe o interesse pelos programas de documentação, principalmente, nas etapas onde métodos e recursos são sistematicamente utilizados na prática da documentação e arquivamento linguístico.

Esses conhecimentos reunidos estimularam reflexões acerca das melhores práticas sobre a utilização eficiente da documentação linguística como instrumento direcionado ao registro multifuncional e exaustivo das línguas ameaçadas, com destaque para indígenas, antes que estas desapareçam sem qualquer tipo de coleta, análise e arquivo de seu corpus documental.

A revisão da literatura indica que essa subárea da Linguística é amplamente

utilizada para a salvaguarda das línguas ameaçadas, contrapondo-se às políticas que contribuem para o aumento do nível de perigo em que se encontram globalmente. Nesse sentido, é importante pensar em um processo de sistematização da documentação linguística que transforme seus procedimentos em prática comum entre os linguistas, desmistificando sua suposta complexidade ao se cogitar sua implementação na pesquisa de campo.

Há muita literatura estrangeira tratando de melhores práticas aplicadas à documentação. Porém, há poucas, ou quase nenhuma referência bibliográfica que descreva todas as etapas desse processo até o arquivamento, publicação e acesso da comunidade ao arquivo, muito menos propondo um arquivo linguístico digital, personalizado, com base nos parâmetros identificados nos arquivos dos programas estudados. Esta dissertação enfrenta esse desafio.

Deste modo, este o estudo justifica-se e mostra-se relevante para a ciência e para os pesquisadores ao apontar, a partir do estado da arte, como as pesquisas da linguística documental são desenvolvidas em etapas que envolvem o emprego de teoria, metodologia, prática, arquivamento e acesso de usuários no âmbito dos programas de documentação.

## 1.6 - Problema

O problema que pretendemos responder com a pesquisa é: “qual é o impacto dos programas DOBES, DMA e AILLA sobre a questão das línguas indígenas ameaçadas?” A intenção é verificar, por meio de uma análise geral dos três programas de documentação, se os recursos empregados (financeiros, acadêmicos, documentais e tecnológicos), em seu conjunto, têm contribuído para o enfrentamento da ameaça à diversidade linguística.

Com esse fito, trataremos da fonte de recursos e dos investimentos feitos pelas instituições, abordaremos as informações sobre o núcleo acadêmico envolvido com o assunto em tela, os principais locais onde as pesquisas são desenvolvidas e o quantitativo de projetos desenvolvidos, bem como aspectos administrativos e técnicos relativos à utilização dos respectivos arquivos linguísticos digitais, entre outras questões.

## 1.7 - Objetivos

### Geral

O objetivo geral da pesquisa é verificar o impacto dos programas DOBES, DMA e AILLA sobre a questão das línguas indígenas ameaçadas a partir de uma abordagem crítica. Para esse propósito, a ideia é avaliar, amplamente, de que forma o conjunto de iniciativas adotadas podem ser validadas ou não como políticas eficientes em torno do objetivo perseguido. Com funções complementares, mas não menos importantes, listamos abaixo os objetivos específicos.

### Específicos

- Identificar, as principais características e ações dos programas DOBES, DMA e AILLA, destacando-os como protagonistas na construção de projetos voltados à documentação das línguas indígenas ameaçadas.
- Identificar entre as diversas características e ações desses programas de documentação, quais são aquelas que precisariam ser reavaliadas visando o aprimoramento.
- Avaliar, entre os três programas de documentação analisados, qual deles apresenta o arquivo linguístico digital com os melhores recursos e determinar as razões pelas quais o melhor de suas características deveria ser replicado e aprimorado.

## 1.8 - Referencial teórico

Em nossa proposta de trabalho consideramos, acerca do referencial teórico, as publicações mais recentes que têm relação com nosso objeto de pesquisa; principalmente as obras que associam o modo pelo qual os programas de documentação utilizam os instrumentos teóricos e metodológicos da documentação linguística para as línguas minoritárias, em especial, as línguas indígenas. Nesse caso, é importante ressaltar que as publicações que abarcam esse espectro de análises são relativamente raras. Os artigos nas linhas de estudos em questão são muitos mais ecléticos na diversificação dos temas e o bom volume de publicações

ajuda aos pesquisadores na tarefa de garimpar fontes tão específicas.

Em torno dessas discussões, acrescenta-se que há desequilíbrio entre as produções acadêmicas (na forma de artigo ou livro). Por exemplo, as publicações direcionadas à documentação linguística e às línguas indígenas ameaçadas superam largamente àquelas preocupadas em investigar as agências de financiamento e os programas de documentação. A maioria das referências que abordam esses eixos de discussão são encontradas, respectivamente, no formato de artigos digitais ou de textos não científicos, como as informações de portais institucionais online (pesquisa documental). Ressaltamos que essas informações foram preciosas, em nosso caso, para investigarmos diversos dados (o Código de Conduta, as informações sobre corpo docente, a história, os objetivos, etc.) dos programas de documentação linguística que são objeto de nossa investigação. No trecho seguinte deste texto destacamos o modo como as principais linhas de discussões bibliográficas, estabelecidas ao longo do tempo, para alguns temas, são publicadas.

Como dissemos, as publicações sobre documentação linguística são muitas e geralmente tratam da diversidade de seus projetos, práticas, estudos e fundamentos. Elas representam o modo pelo qual a área vem se consolidando e crescendo. Ilustrativamente, destacamos algumas dessas obras: Documentação e descrição da língua<sup>25</sup>, publicada desde 2003 pela *School of Oriental and African Studies* (SOAS), da Universidade de Londres – artigos revisados por pares sobre documentação linguística resultado de workshops realizados na SOAS; Manual Cambridge de línguas em extinção<sup>26</sup>, de Austin & Sallabank, publicado em 2011 – tem importância conceitual e introdutória, indicando o papel dos falantes, os tipos de dados e estruturas, como arquivar digitalmente e gerenciar projetos; Documentação da língua: prática e valores<sup>27</sup>, de Grenoble & Furbee, publicada em 2010 – reúne uma série de experiências em trabalho de campo mediante estudos de caso que destacam práticas e valores, medidas de adequação, uso da tecnologia e técnicas colaborativas; Palestras em documentação e descrição de línguas<sup>28</sup>, de Austin, publicada em 2010

---

<sup>25</sup> Cf. Título e referência original: Language Documentation and Description. 2003–2012. London: School of Oriental and African Studies, Univ. of London.

<sup>26</sup> Cf. Título e referência original: Austin, Peter K., and Julia Sallabank, eds. 2011. *The Cambridge handbook of endangered languages*. Cambridge, UK: Cambridge Univ. Press.

<sup>27</sup> Cf. Título e referência original: Grenoble, Lenore A., and N. Louanna Furbee, eds. 2010. *Language documentation: Practice and values*. Amsterdam: John Benjamins.

<sup>28</sup> Cf. Título e referência original: Austin, Peter K. Lectures in language documentation and description. Language Documentation and Description 7. London: School of Oriental and African Studies, Univ. of London, 2010.

– trata de problemas e métodos de documentação, arquivamento, gravação de áudio, linguagem de sinais, ética, política linguística, tipologia, teoria linguística e acesso a bolsa de pesquisa; Revisão dos Essenciais da documentação linguística<sup>29</sup>, de Evans, publicada em 2008 - dialoga com Gippert (2006) ao direcionar sua revisão crítica a pouca inserção da escrita gramatical na documentação linguística; Essenciais da documentação linguística<sup>30</sup>, de Gippert, Jost & Himmelmann, publicada em 2006 – um dos poucos trabalhos de referência acerca dos princípios teóricos e práticos da documentação linguística, a obra segue a influência e o modelo desenvolvido pelo Programa DOBES. Há uma versão em espanhol do mesmo livro, intitulado “Bases de la documentación lingüística”<sup>31</sup>, publicada em 2007. Há centenas de outras obras, algumas das quais podem ser consultadas nas referências desta dissertação.

### 1.9 - Metodologia

A metodologia é o estudo dos métodos reconhecidos pela ciência, com destaque para aqueles mais adequados à execução de uma pesquisa. O método científico é reconhecido pela comunidade acadêmica, em relação aos procedimentos, por aproximar-se da exatidão nos resultados na busca por determinados objetivos.

Sobre essa questão, Lakatos e Marconi (2003, p. 83) afirmam que “o método é o conjunto das atividades sistemáticas e racionais que, com maior segurança e economia, permite alcançar o objetivo – conhecimentos válidos e verdadeiros -, traçando o caminho a ser seguido, detectando erros e auxiliando as decisões do cientista”.

---

<sup>29</sup> Cf. Título e referência original: Evans, Nicholas. Review of *Essentials of Language Documentation*. *Language Documentation and Conservation*, 2008. p. 340–350.

<sup>30</sup> Cf. Título e referência original: Gippert, Jost, Nikolaus P. Himmelmann, and Ulrike Mosel, eds. 2006. *Essentials of language documentation*. Trends in Linguistics, Studies and Monographs 178. Berlin: Mouton de Gruyter.

<sup>31</sup> Cf. Título e referência original: Flores, Farfán J. A, and John B. Haviland. *Bases De La Documentación Lingüística*. México, DF: Instituto Nacional de Lenguas Indígenas, 2007.

Assim, apresentamos os procedimentos a serem seguidos na realização da investigação. Sua organização foi estruturada e definida conforme as características desta dissertação. Na sistematização da metodologia científica adotada optou-se por sua divisão em elementos essenciais, divididos em categorias, conforme o que segue:

- A finalidade da investigação é a pesquisa aplicada porque partimos de conhecimentos básicos produzidos em pesquisas precedentes e semelhantes como ponto de partida para resolver o problema proposto nesta dissertação. Não se objetiva somente produzir conhecimento teórico útil, pretende-se, sobretudo, que este conhecimento sistematizado seja aplicado em todas as etapas da documentação linguística, inclusive na pesquisa de campo.
- No que se refere à abordagem, escolheu-se a pesquisa quali-quantitativa porque ambas as faces da pesquisa se complementam em torno dos objetivos da análise e o investigador pode selecionar os atributos que julgar mais importantes. A implementação desse enfoque dual considera a priorização do aspecto puramente teórico da pesquisa bibliográfica e enfatiza as características técnicas e matemáticas específicas da tecnologia da informação.
- Em relação ao método, definiu-se o hipotético-dedutivo, por tratar-se de uma abordagem que funciona a partir da proposição de um problema. A observação do objeto de estudo deve permitir ao pesquisador identificar hipóteses, testá-las, descartar as reprovadas no teste e tirar conclusões sobre o problema.

Quanto aos procedimentos, foram selecionados os dois destacados abaixo:

- A pesquisa bibliográfica, por tratar-se do procedimento mais utilizado para a investigação realizada, utilizando-se como fontes livros, artigos e textos científicos publicados. Esse tipo de pesquisa tem caráter teórico e investiga a relação dos conceitos, ideias e características do objeto de modo que a comparação de várias opiniões sobre um problema alimente o pesquisador de informações sobre o assunto. A posse dessas informações deve possibilitar-lhe a elaboração de comparações e reinterpretações críticas daquilo que já conhece, no nível teórico, com os novos conhecimentos que está adquirindo e acumulando.

- Por fim, a utilização da pesquisa documental se justifica porque serão utilizados textos não científicos na pesquisa essencialmente teórica.

Encerrada essa breve introdução, convido o leitor a ingressar nessa caminhada em busca dos fatores que levaram os programas de documentação a se destacarem no cenário internacional de ameaça à diversidade linguística ao longo dos últimos vinte e cinco anos. Para começarmos a percorrer esse labirinto, iniciaremos pelos elementos que desencadearam o interesse por essa ainda recente abordagem documental.

## 2 – REVISÃO DA LITERATURA

### 2.1 – Línguas indígenas ameaçadas

#### 2.1.1 - Ameaça em números

A maioria dos programas e projetos direcionados às línguas ameaçadas surgidos no início do século XXI, no contexto das preocupações aqui narradas, continuam existindo e exigem, diante das constantes ameaças às línguas minoritárias, que novas iniciativas protetivas sejam construídas e apoiadas. Assim, são louváveis, conforme destacam Austin & Sallabank (2011, p. 3), os esforços conjuntos das agências intergovernamentais preocupadas com a perda da diversidade linguística acompanhados de ações, como as 4 escalas que medem o nível de ameaça linguística, com destaque para as categorizações da(o): Graded Intergenerational Disruption Scale (GIDS); UNESCO; Ethnologue; e Expanding Graded Intergenerational Disruption Scale (EGIDS)<sup>32</sup>, resumidamente discutidas abaixo:

A escala GIDS, por duas décadas referência no estudo da ameaça à diversidade linguística, forneceu os fundamentos teóricos para a revitalização da língua. Resultado dos esforços de Fishman (1991), relaciona o shift linguístico à perda das funções desempenhadas por determinada língua na sociedade até a sua (possível) morte. Essa escala é orientada pelas funções ou domínios sociais da língua, elementos que lhe permitem organizar informações sobre a ameaça linguística e, conseqüentemente, políticas voltadas à perda dessas funções (para maiores detalhes sobre a escala, ver FISHMAN, 1991; SIMONS & LEWIS, 2013).

A UNESCO desenvolveu a sua escala de 6 níveis para avaliar a situação da vitalidade das línguas ameaçadas, conforme proposição de alguns linguistas. Em seu quadro, fixa seis categorias em escala de vitalidade das línguas associadas ao conjunto de nove fatores cuja análise determina a categoria. O mais importante desses fatores é a transmissão intergeracional (para maiores detalhes sobre a escala, ver BREZINGER et al., 2003; SIMONS & LEWIS, 2013).

O Ethnologue representa outro esforço de categorização da vitalização linguística em escala de cinco níveis com ênfase no número de falantes da primeira língua em detrimento de outros fatores (para maiores detalhes sobre a escala, ver GORDON, 2005; GRIMES, 2000; LEWIS, 2009; SIMONS & LEWIS, 2013); para maior

---

<sup>32</sup> Confira as escalas em Simons & Lewis (2013, p. 1-18).

detalhamento das observações críticas que levaram à elaboração de uma nova escala para aferir a vitalidade das línguas (EGIDS), ver Simon & Lewis (2013).

A escala EGIDS ao combinar categorias das escalas citadas resulta em escala de treze níveis que visa analisar a perturbação da transmissão intergeracional. A categorização dessa escala permite diferenciar “língua extinta” de “língua dormente” e, por conseguinte, determinar a necessidade ou não de revitalização linguística. Entre as principais características dessa escala, destacam-se:

- As análises baseadas em escala de avaliação global dos níveis de ameaça que permitiram a formulação de novas conclusões sobre as previsões de 1992 (ver KRAUSS, HALE, etc.). Para mais informações ver Simons & Lewis (2013).
- Os critérios dessa escala, para qualquer língua conhecida, até para as que não possuem falantes, podem ser categorizados. A avaliação da língua ocorre conforme as respostas às cinco questões-chave relativas às funções de identidade, veicularidade, estado da transmissibilidade entre gerações (item importante para as iniciativas de documentação), estado de aquisição de alfabetização e o perfil social de uso geracional da língua (SIMONS & LEWIS, 2013).

Assim, a EGIDS é uma escala aperfeiçoada, elaborada a partir das críticas feitas por Simons e Lewis (2013) às três escalas mencionadas. Contudo, os autores não desprezam esses instrumentos e continuam a acentuar a importância que eles possuem no universo dos estudos relacionados à ameaça linguística. Abaixo, segue o Quadro I com a escala EGIDS.

## GIDS expandida

Escala de Interrupção Intergeracional Graduada Expandida (adaptado de Fishman 1991) <sup>33</sup>			
NÍVEL	ATRIBUTO	DESCRIÇÃO	UNESCO
0	Internacional	A língua é utilizada internacionalmente para uma vasta gama de funções.	Seguro
1	Nacional	A língua é utilizada na educação, no trabalho, nos meios de comunicação social, no governo a nível nacional.	Seguro
2	Regional	A língua é utilizada para os meios de comunicação social locais e regionais e serviços governamentais.	Seguro
3	Comércio	A língua é utilizada para trabalhos locais e regionais, tanto por pessoas de dentro como por pessoas de fora.	Seguro
4	Educação	A alfabetização na língua é transmitida através de um sistema de educação pública.	Seguro
5	Escrita	A língua é utilizada oralmente por todas as gerações e é efetivamente utilizada na forma escrita em partes da comunidade.	Seguro
6a	Vigorosa	A língua é usada oralmente por todas as gerações e é aprendida pelas crianças como a sua primeira língua.	Seguro
6b	Ameaçada	A língua é utilizada oralmente por todas as gerações, mas apenas algumas das gerações as transmitem aos seus filhos.	Vulnerável
7	Mudança	A geração de filhos conhece suficientemente bem a língua para a utilizar entre si, mas nenhum deles a transmite aos seus filhos.	Definitivamente em perigo
8a	Moribunda	Os únicos falantes ativos da língua que restam são membros da geração dos avós.	Severamente Ameaçada
8b	Quase extinta	Os únicos falantes restantes da língua são membros da geração dos avós ou mais velhos que têm poucas oportunidades de usar a língua.	Criticamente Ameaçada
9	Dormente	A língua serve como lembrete de identidade patrimonial para uma comunidade étnica. Ninguém tem mais do que proficiência simbólica.	Extinta
10	Extinta	Ninguém retém um sentido de identidade étnica associada à língua, mesmo para fins simbólicos.	Extinta

Quadro I: Escala EGIDS – análise da perturbação intergeracional<sup>33</sup>

Para mais detalhes sobre os níveis descritos acima, veja o resumo em Simons & Lewis (2013, p. 111-113).

Como vimos, o conhecimento do nível de transmissão intergeracional da língua é um elemento fundamental para que lhe seja atribuída um determinado grau de vitalidade. A complexidade dos estudos relacionados à questão da ameaça às línguas minoritárias provoca o surgimento de novas metodologias visando tornar mais precisa a aferição do grau de perigo à diversidade linguística. Porém, como dissemos, a ampliação da aceitação de escalas como a EGIDS não impede os pesquisadores de

<sup>33</sup> Quadro retirado e adaptado de Simons & Lewis (2013, p. 110).

estudar outras escalas, como o Ethnologue, em torno de diversos temas, como as línguas minoritárias.

As interpretações gerais sobre as línguas minoritárias são extraídas de listas estatísticas que apresentam dados relativos ao número de línguas faladas, com destaque para aquelas que sofrem o risco de desaparecer. Segundo Austin & Sallabank (op. cit., p. 3), a fonte estatística mais utilizada para colher informações sobre o número de línguas e respectivos usuários é o *Ethnologue Languages of the World*<sup>34</sup> que destaca a existência de mais de 7000 línguas vivas em seu endereço eletrônico<sup>35</sup>. Porém, há críticas quanto aos procedimentos para a contagem das línguas. Alguns desses críticos, como Moore (2011) apud Storto (2019, p. 11), usam o critério da “inteligibilidade mútua” para diferenciar as línguas a partir do seguinte princípio: se falantes de duas variedades linguísticas não conseguem se entender, as variedades são consideradas línguas distintas. Se eles puderem se entender, as variedades são consideradas variedades geográficas da mesma língua. Segue abaixo um gráfico Ethnologue com os números percentuais de concentração das línguas por continente.

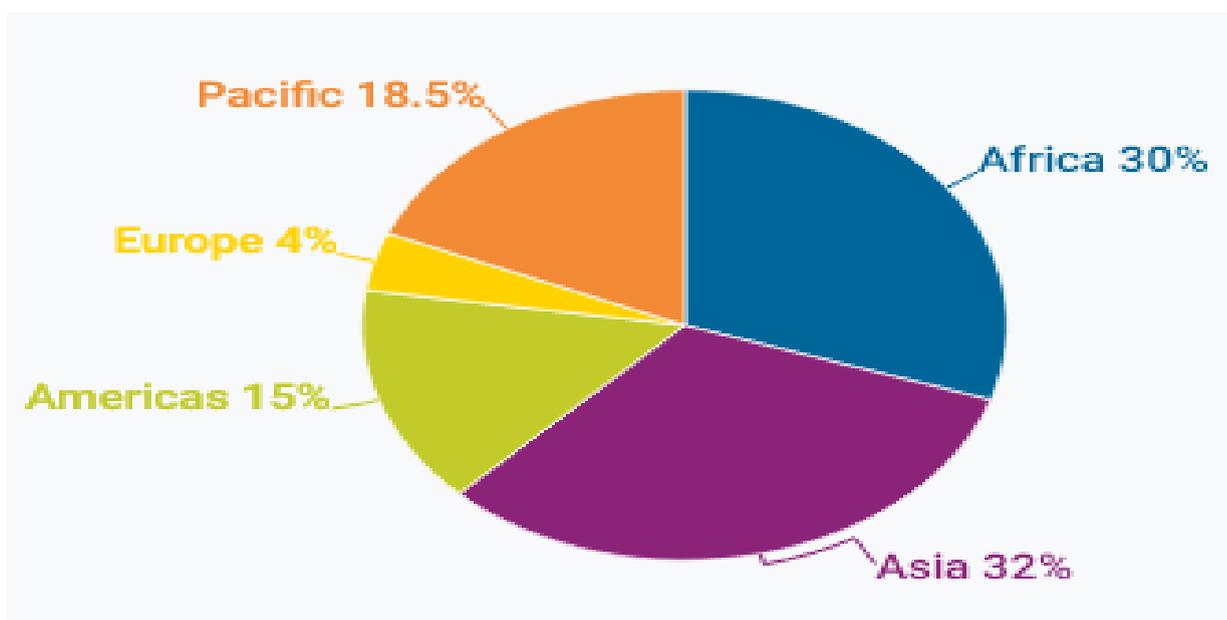


Gráfico I: Percentual de concentração das línguas por continente<sup>36</sup>

<sup>34</sup> Cf. Ethnologue. Disponível em: <<https://www.ethnologue.com/about/history-ethnologue#:~:text=The%20Ethnologue%20was%20founded%20by,as%20with%20other%20language%20researchers>>. Acesso em: 28 set. 2020.

<sup>35</sup> Cf. Guides Ethnologue. Disponível em: <<https://www.ethnologue.com/guides/how-many-languages>>. Acesso em: 30 set. 2020.

<sup>36</sup> Gráfico retirado e adaptado de Ethnologue. Disponível em: <<https://www.ethnologue.com/guides/continents-most-indigenous-languages>>. Acesso em: 30 set. 2020.

Em linha com as reclamações contra o modelo de apuração de dados estatísticos do *Ethnologue*, os críticos abordam a defasagem das informações baseadas em estimativas de 1976, enquanto os dados mais recentes do censo de 2001 são ignoradas. Os próprios editores do documento reconhecem sua incapacidade de acompanhar a evolução dos dados relativos às línguas. Por isso, os pesquisadores e a UNESCO continuam a utilizar o “Atlas of the world's languages in danger” (MOSELEY, 2012), obra que para muitos, apesar das deficiências, ainda é um guia útil na apresentação de dados relativos à diversidade linguística em todo o mundo.

Austin & Sallabank (2011, p. 3) pontuam que “[...] a inteligibilidade mútua pode ocorrer, em certa medida, em função das atitudes e da política; isto é, se as pessoas querem ou não se entender. Tais atitudes estão, em certa medida, ligadas à questão de uma comunidade se considerar parte de uma determinada identidade étnica e linguística distinta, mas alguns membros dessa comunidade não concordarem com isso [...]”. O aspecto relacionado à diferenciação das línguas, a partir de princípios políticos, vem sendo defendido desde o século XIX por filósofos que veem a língua como um instrumento fundamental da identidade através do dístico "um estado, um povo, uma língua". Esse pensamento é visto como um ideal em contexto altamente nacionalista que não raramente provoca ações violentas, de características etnocidas, aos falantes de línguas minoritárias.

Um caso importante na relação entre língua e política ocorre no caso do Quechua, visto por alguns pesquisadores como família linguística e por outros como uma língua, mas como afirmam Coronel-Molina & McCarty (2004) apud Austin & Sallabank (2011, p. 4) este é, na verdade: “[...] um nome abrangente que denota um grupo de variedades linguísticas relacionadas [...]”. Apesar dos linguistas destacarem vinte e sete línguas quechuas no Peru, o governo daquele país reconhece somente seis como línguas. Em torno dessa questão, é importante destacar a manipulação estratégica de grupos separatistas que enfatizam as diferenças linguísticas para angariar apoio em torno de sua causa e de acordo com seus interesses, enquanto os governos, por outro lado, procuram ignorá-las. Na página supracitada, Austin & Sallabank (2011) apontam que os “[...] paradoxos como a incompreensibilidade mútua das variedades linguísticas chinesas em comparação com a compreensibilidade mútua das línguas escandinavas do continente são claramente motivados por

considerações políticas e nacionalistas, e não linguísticas."

A divergência em relação aos critérios de uma contabilização factível do número de línguas existentes no mundo feita pelo Ethnologue ou pelo *Atlas of the world's languages in danger* (UNESCO) aponta para a complexidade desse tipo de levantamento que impede, a partir dos critérios de cada estudioso do assunto, o uso de uma metodologia ortodoxa confiável para esse tipo de levantamento.

Esse desacordo levou a busca de novos caminhos que norteiem as pesquisas dos linguistas em torno da questão das línguas, de modo geral, e das línguas minoritárias ameaçadas, em particular. Com base nesses elementos, sociolinguistas e linguistas antropológicos, utilizando teorias mais modernas acerca desse assunto, questionam se é realmente possível fazer uma contagem precisa do número de línguas existentes, pois fatores étnicos, políticos e de inteligibilidade mútua dificultam esse esforço.

### 2.1.2 – Fatores de ameaça

No levantamento voltado à definição dos fatores determinantes para a classificação de "línguas ameaçadas", pesquisadores como Nettle & Romaine (2000), Crystal (2000) e Grenoble (2011) apud Austin & Sallabank (2011, p. 5), identificaram e resumiram a partir de suas obras as seguintes categorias:

- catástrofes naturais, fome, doença, por exemplo, Malol, Papua-Nova Guiné (terremoto); Ilhas Andaman (tsunami);
  - guerra e genocídio, por exemplo, Tasmânia (genocídio dos colonos); povos indígenas brasileiros (disputas por terra e recursos); El Salvador (guerra civil); repressão aberta, geralmente em nome da "unidade nacional" ou assimilação, incluindo reassentamento forçado: por exemplo, idiomas curdo, galês e nativo americano;
  - domínio cultural / político / econômico, por exemplo, Ainu, Manx, Sorábio, Quechua e muitos outros. Os fatores geralmente se sobrepõem ou ocorrem juntos. As linhas divisórias podem diferir para distingui-los. Por exemplo, nas Américas e na Austrália, a doença e a supressão das culturas indígenas se espalharam após a colonização, e na Irlanda muitos falantes de irlandês morreram ou emigraram devido à inação do governo colonial que agravou os efeitos da fome da praga da batata no século XIX.
- A quarta categoria, considerada "mais comum", foi subdividida em cinco fatores:
- econômico, por exemplo, a pobreza rural leva à migração para as cidades e locais mais distantes. Se a economia local melhorar, o turismo pode trazer falantes das línguas majoritárias;

- domínio cultural pela comunidade majoritária, por exemplo, educação e literatura apenas através da língua majoritária ou da língua oficial do estado. A língua e a cultura indígenas podem ser “folclorizadas”;
- político: políticas educacionais que ignoram ou excluem línguas locais, falta de reconhecimento ou representação política, proíbe o uso de línguas minoritárias na vida pública;
- histórico: colonização, disputas de fronteira, ascensão de um grupo e sua variedade de línguas ao domínio político e cultural;
- atitudinal: línguas minoritárias se associam com pobreza, analfabetismo e dificuldades, enquanto a língua dominante está associada ao progresso.

A compilação de fatores que determinam o quadro preocupante das línguas ameaçadas no mundo expõe uma situação que define o grau de contribuição do homem para o agravamento do problema. Excetuando-se as questões de ordem natural, os demais elementos presentes no levantamento (econômicos, culturais, políticos, históricos e comportamentais) demonstram o envolvimento direto e/ou indireto da ação humana que, ao longo do tempo, tem contribuído, e muito, para aprofundar o atual quadro de ameaças sobre as línguas minoritárias.

### 2.1.3 - Maiores riscos

Outro olhar sobre o fenômeno em que línguas desaparecem sendo substituídas por outras, conforme os linguistas, ocorre a partir do contato linguístico, algo que é constitutivo da história e até foi recentemente visto como um ciclo normal de mudança (DALBY, 2002). Atualmente há consenso de que as línguas não mais aprendidas pelas crianças juntamente com a tendência de sua perda (por exemplo, devido à adoção pelos falantes indígenas das línguas hegemônicas) representa um desafio maior para os pesquisadores que precisam ser ágeis na documentação dessas línguas ameaçadas a fim de que futuras gerações possam conhecê-las e talvez até revitalizá-las. Essa situação atinge um espectro crescente de línguas que possuem um número ínfimo de falantes.

Ao definir Diversidade Cultural, em seu endereço na Internet, a UNESCO<sup>37</sup>, em linhas gerais, diz que esta não deve ser associada somente à produção material, mas a uma forma de se atingir uma vida mais gratificante. Nesse caso, a diversidade linguística é vista como um " pilar da diversidade cultural", pois as línguas por trazerem

---

<sup>37</sup> Cf. UNESCO. Disponível em: <<http://www.unesco.org>>. Acesso em: 05 set. 2020.

elementos fundamentais na relação do homem com a sociedade, como "a identidade, comunicação, integração social, educação e desenvolvimento, têm grande importância para as pessoas e o planeta". Assim, quando as línguas perecem, desaparece essa vasta produção imaterial presente na diversidade cultural que envolve tradições, memória, maneiras peculiares de pensamento e expressão das ideias.

Em sintonia com esse pensamento, David Harrison (2007), na obra *When Languages Die*, demonstra de que modo determinados aspectos das culturas, das sociedades e das línguas humanas estão entrelaçados sendo afetados mutuamente. Nesses casos, a cultura oral possui tradições culturais expressas pela língua por intermédio da linguagem em forma de histórias, lendas, narrativas, poesia e canções. Harrison defende que quando uma língua ameaçada é definitivamente perdida, perde-se também o conhecimento e a riqueza cultural para a comunidade falante e para a humanidade em geral. Para a UNESCO esse conhecimento imaterial pode ser definido também como "patrimônio cultural intangível".

Conforme Austin & Sallabank (2011, p. 7), alguns estudiosos associam e correlacionam os conceitos de diversidade cultural e linguística com diversidade biológica. Pesquisadores, destacadamente Sutherland (2003), relacionaram na Indonésia e Papua-Nova Guiné, a grande variedade de espécies animais com a grande variedade de línguas naqueles locais. Tal correlação, se considerada em termos de diversidade linguística, seria inversamente proporcional à situação da Europa. Essa abordagem, na época considerada inovadora, chamou a atenção de instituições de pesquisa, de pesquisadores e ativistas por negar a mesma relação de causalidade entre o desaparecimento de espécies animais e o desaparecimento das línguas.

Essas discussões, refletem, na década de 1990, a definição de Haugen (1972, p. 325) sobre a *language ecology*: "é o estudo das interações entre qualquer língua dada e seu meio ambiente", sendo que "o verdadeiro ambiente da língua é a sociedade que a usa como um de seus códigos". Naquele momento, linguistas, de diferentes linhas, preocupavam-se com as ameaças à diversidade linguística e ao meio ambiente; concomitantemente, a ecolinguística despontava como nova disciplina. Inicialmente, isso provocou confusões conceituais entre os muitos linguistas que acreditavam existir relação entre a nova teoria e movimentos ambientalistas. De caráter holístico e abrangente sobre fenômenos linguísticos, a nova especialidade

científica possui afinidades temáticas com a linguística ambiental e a linguística crítica, o que possibilita absorvê-las (COUTO, 2013).

Outra linha interpretativa dessa questão possui viés político e associa o declínio da diversidade linguística e biológica às consequências da globalização e/ou do capitalismo internacional. Segundo Austin & Sallabank (2011, p. 8), os ecolinguistas costumam relacionar a língua com um organismo vivo, o que Mackey (2001) trata como uma falácia. Segundo ele, como as línguas são artefatos humanos, não espécies, elas não possuem vida própria fora dessas comunidades. Portanto, as comunidades humanas precisam ser sustentáveis para manter suas línguas.

Apesar de não termos como pretensão avançar nessa relação que alguns pesquisadores fazem com a noção de diversidade, tanto no âmbito das línguas quanto na seara da natureza animal, esses pontos de contato parecem que ao longo do tempo suscitaram mais divergências do que consenso. Tudo levar a crer que as correntes de estudo da diversidade linguística são aquelas que reservam uma área exclusiva para esse tipo de análise, ou seja, aquela que não imiscua conceitos de outras áreas de conhecimento. A interdisciplinaridade é importante, mas deve ser implementada de modo que conteúdos específicos não se confrontem, dificultando uma análise mais acurada do que se cogita investigar.

Uma parcela de linguistas entende que as línguas representam símbolos de identidade étnica e nacional. Em torno dessa ideia, aqueles envolvidos na proteção às línguas ameaçadas acreditam que quando uma língua morre, um modo particular de interpretar o mundo também desaparece<sup>38</sup>. Essa corrente de pensamento concorda com a hipótese da relatividade linguística de Edward Sapir (1949) e Benjamin Lee Whorf (1956)<sup>39</sup>, segundo a qual o modo peculiar que a língua possui para decifrar a realidade demonstra a importância da linguagem para a ordenação do pensamento e formação das singularidades nas concepções de mundo. Deste modo, a maneira como interpretamos a realidade resulta dos hábitos de linguagem construídos socialmente (MARTELOTTA, 2011, p. 55). Há autores pós-modernistas que rejeitam a ideia de identidades étnicas e nacionais formais e fixas ao defender a interpretação de que elas devem ser encaradas como fluidas, construídas frequentemente consoante o contexto social cotidiano das pessoas (CASTELLS, 2004; OMONIYI &

---

<sup>38</sup> Cf. Dalby (2002), Fishman (1988), Nettle & Romaine (2000), Storto (2019).

<sup>39</sup> Cf. ALVES, Leonardo M. Hipótese Sapir-Whorf. Disponível em: <https://Hipótese Sapir-Whorf – Ensaios e Notas>> Acesso em: 08 set. 2020.

WHITE, 2008). Apesar das discordâncias existentes, Austin & Sallabank (2011) acreditam que a manutenção da identidade regional é um elemento importante no contexto da globalização. Em consonância com isso, autores como Lanza & Svendsen (2007) entendem que defender a língua é fundamental para a manutenção da identidade quando determinado grupo social, por questões de caráter político e social, a tem ameaçada.

Podemos perceber que os referenciais teóricos que tratam da identidade étnica têm sofrido mudanças ao longo do tempo. Certamente essa variação de objetivos acompanha o contexto político, econômico e social no espaço internacional em que são produzidos. As reflexões que associam língua e nacionalidade têm relação direta com o período em que essa ideologia prevaleceu. Por outro lado, as interpretações de caráter mais científico que procuram isolar a identidade étnica de conceitos políticos envolvem a preocupação com o aspecto puramente científico da Linguística. As análises que ressaltam a importância da “fluidez” da noção desse mesmo conceito têm relação direta com a pluralidade e a interseção cultural típica do globalismo. E, por fim, a defesa das diferenças através da manutenção das particularidades culturais, ao nível mundial, permite ressaltar a riqueza que a diversidade representa.

Como vimos, as últimas décadas marcam a conscientização mundial de que a diversidade linguística está ameaçada e, por consequência, se alastra a diminuição de línguas e de falantes das línguas minoritárias. Como foi discutido, vários linguistas se posicionaram sobre essa questão. Diante da gravidade do problema, a maioria deles acredita que a perda da diversidade linguística é um fenômeno extremamente grave, de modo que é fundamental que eles aprendam, debatam mais sobre o assunto e proponham saídas para o caso. Um ponto defendido é a necessidade de se documentar as línguas ameaçadas antes que elas desapareçam. Há posicionamentos distintos em relação a esse assunto. Há os que acreditam (MAFFI, 2003) que o processo acelerado de documentação, em frágeis bases teórico-metodológicas envolvendo a coleta de material, pode implicar na qualidade final dos arquivos de dados reunidos, gerando como resultado a manipulação inadequada da língua. Contudo, a realidade tem mostrado, claramente, que os objetivos e fundamentos da documentação linguística possibilitam uma melhor compreensão do fenômeno ao apresentar respostas positivas para esse problema global, e dar margem para que amplos seguimentos de estudiosos do assunto possam aplicar seus conhecimentos sobre os arquivos documentais digitalizados.

## 2.2. - Línguas indígenas na América do Sul

### 2.2.1 – Paradigma da diversidade

Contraopondo-se ao quadro sombrio que se delineia para as línguas minoritárias, Moore (2007) destaca o papel relevante das línguas indígenas da América do Sul no cenário mundial. Segundo ele, são cerca de 90 famílias linguísticas e um número indeterminado de línguas isoladas que superam, e muito, as 14 existentes no continente africano. A pesquisa de Baéz, Rogers & Labrada (2016, p. 2), por outro lado, reforça esses números e acrescenta informações ao afirmar que “[...] a América Latina abriga mais de 100 famílias de línguas distintas, faladas em comunidades de todos os tamanhos (variando de milhões de falantes, em alguns locais, a um único falante, em outro). Algumas dessas línguas tornaram-se línguas oficiais, como o Guaraní, Quechua e o Aymara, enquanto as demais sequer são reconhecidas por seus governos nacionais [...]”. Sobre os aspectos demográficos, em abordagem histórica, Lizarralde (2001) apud Moore (2007, p. 29) estima que havia “possivelmente 1200 grupos indígenas” na América do Sul antes do contato com os europeus, e que 65% das línguas nativas foram extintas; enquanto Kaufman e Golla (2001) apud Moore (2007, p. 29) calculam 550 línguas nativas na América do Sul pré-colombiana, das quais 300 sobrevivem. Em torno da presença de línguas isoladas nesse território, os mesmos estudiosos estimam em 50, o número de famílias e línguas isoladas na região.

Diante desse quadro difícil para as línguas minoritárias no mundo, os linguistas que se debruçam sobre a questão das línguas indígenas sul-americanas costumam identificar, além da diversidade linguística, as grandes variações no quantitativo das populações ameríndias com impacto sobre o número de falantes. É importante dizer que as projeções numéricas sobre a realidade dos povos indígenas expostas, fazem parte de estudos contínuos conduzidos por diferentes pesquisadores que possuem distintos objetivos com seus trabalhos. Portanto, essas quantificações representam isso. Os dados seguintes apontam algumas informações relevantes sobre as maiores populações indígenas sul-americanas e as suas localizações.

### 2.2.2 – Fatos e números essenciais

Segundo Grinevald (1998), as línguas indígenas da América do Sul estão geograficamente difusas e não respeitam fronteiras. Isso acontece com aquelas faladas em diversos países, entre as quais destacam-se: o Quechua, em 5 países; o Aymara, em 3 países; o Guaraní, em 3 países; e o Mapuche, falado em 2 países. O mesmo espalhamento ocorre com as línguas menores, onde os falantes sobreviventes são encontrados nas fronteiras dos nove países da bacia amazônica. Conforme Moore (2007), essas línguas e as demais que serão analisadas, são encontradas em uma vasta região que corresponde aproximadamente às terras baixas da América do Sul: as regiões amazônicas da Bolívia, Peru, Equador e Colômbia, assim como todo o Brasil, Venezuela, Guiana, Suriname e Guiana Francesa. Os números atuais indicam a presença de cerca de 300 línguas indígenas e crioulas nessa região, embora exista a possibilidade desse número ser menor, no caso das variedades linguísticas mutuamente inteligíveis não serem contabilizadas como línguas.

A questão da inexistência de uma fronteira rígida para as línguas indígenas sul-americanas representa um desafio para os linguistas no sentido de se compreender de que modo elas conseguiram manter, geneticamente, esse grau de diversidade, bem como mantiveram e mantêm um contato tão representativo entre si, nessa região. Afinal, o que resultou desse espraiamento das línguas indígenas sul-americanas? Restaram mais línguas (no sentido de variedade linguisticamente independente e distinta de outras línguas) ou ficaram simplesmente as variedades linguísticas oriundas desses contatos? Essas são questões que os linguistas têm estudado e procurado responder.

Em outra abordagem sobre o quantitativo das línguas indígenas existentes nessa região, Crevels (2012) apresenta o quadro abaixo que representa algo bastante ilustrativo no que tange à população aproximada de indígenas na América do Sul, tomando-se os devidos cuidados com a coluna "população indígena" que, segundo ele, pode apresentar números inconsistentes devido à subnotificação. Contudo, seria interessante analisar, no caso do Brasil, que 1) apesar da ínfima população indígena, o país possui o maior número de línguas indígenas e, por fim, 2) na relação entre a população total e a população indígena, o percentual de participação da última é quase desprezível: 0,2%.

<b>País</b>	<b>População total</b>	<b>População indígena</b>	<b>Línguas indígenas</b>
Brasil	198,739,300	358,000	177
Colômbia	45,644,000	1,392,600	68
Peru	29,547,000	4,045,700	50
Venezuela	26,814,800	536,900	36
Bolívia	9,827,500	4,541,000	33
Paraguai	6,995,700	103,300	18
Argentina	40,913,600	600,300	15
Equador	14,573,100	830,400	13
Guiana	772,300	50,000	10
Suriname	481,300	7,000	8
Guiana Francesa	221,500	5,000	7
Chile	16,601,700	692,200	6
<b>TOTAL</b>	<b>391,131,800</b>	<b>13,162,400</b>	<b>420</b>

Quadro II: Populações indígenas e número de línguas indígenas na América do Sul<sup>40</sup>

A parte do território sul-americano que não abrange o Brasil, abriga as línguas ameríndias com mais de 100.000 falantes. Nesse quadro, o Quechua deve ser destacado por ser a língua mais falada na região, com valores que oscilam de 8 a 12 milhões<sup>41</sup>, de modo que os falantes se distribuem da seguinte forma: Equador, 2 milhões; Peru, 4,5 milhões; e Bolívia, 1,5 milhões. Regionalmente, a língua tem importância fundamental entre as demais por ser ensinada como segunda língua em locais como universidades, missões religiosas, etc. Conforme Adelaar (1991, p. 77), a oficialização da língua, ocorrida em 1975, foi acompanhada da tentativa política de reconhecer ao menos seis das principais variedades linguísticas e lhes dar a mesma importância, ao invés de escolher somente uma dessas variações. Porém, decidiu-se pela escolha de um alfabeto oficial de Quechua a partir do qual todas as outras variedades linguísticas poderiam ser escritas. Pode-se citar também o impacto educacional em algumas sociedades, pois a língua é usada no processo de educação bilíngue que se transformou em programa regular de educação nos três países andinos onde é falada (GRINEVALD 1998, p. 130).

Em relação ao Aymara, a população gira em torno de 2 a 3 milhões de falantes, cuja maioria se localiza na Bolívia; é também a língua das pessoas que habitam o norte do Chile e o sul do Peru. A língua tem origem no altiplano Collao, área que rodeia o Lago Titicaca. Conforme Briggs (1985) apud Grinevald (1998, p. 130), em números aproximados, há 350.000 falantes no Peru e 25.000 no Chile. A língua funciona há

<sup>40</sup> Quadro retirado e adaptado de Crevels (2012, p. 167).

<sup>41</sup> Para mais dados sobre falantes do Quechua consulte Kaufman (1994), Parker (1969), Cerrón-Palomino (1987) e Adelaar (1991).

tempos como língua franca e faz parte de programas de padronização linguística, de educação bilíngue e de programas universitários de treinamento de professores na Bolívia e Chile.

A família Maia representa o terceiro conjunto de línguas mais faladas na América Latina, e compõe as línguas nativas de populações que residem na Guatemala e México. Outra família linguística dessa região é a Uto-Azteca (parte do ramo Azteca) da qual faz parte a língua Náhuatl, falada por aproximadamente dois milhões de indígenas localizados no sudeste do México (DUARTE, 2016). Porém, há variedades linguísticas em outras partes do país, como no Noroeste; além da língua Pipil que lhe é similar, sendo falada em ambos os lados da fronteira entre Guatemala e El Salvador.

Segundo Grinevald (1998, p. 131), o Mapudungu ou Mapudungun, língua do povo mapuche, é conhecida também como arauciano ou mapuche. É usada principalmente no Chile, onde possui um número de falantes que varia de 200.000 a 500.000. Se contabilizada a população que vive na capital, cerca de 300.000 pessoas, esse número pode chegar a um milhão. Politicamente, esse grupo étnico atua por questões relacionadas a direitos à terra e à manutenção e desenvolvimento de sua língua.

Por fim, destacamos o Guaraní. A língua nativa é falada por dez milhões de pessoas no Paraguai, e em regiões do nordeste da Argentina, sul e sudeste do Brasil e Bolívia (DUARTE, 2016, p. 32). A maioria dos falantes é composta por mestiços que falam o espanhol e não por indígenas. Apesar do longo tempo reconhecida como língua nacional, ou seja, língua da maior parte da população, somente em 1992 foi aceita como língua do Paraguai. Com isso, a Constituição de 1994 chancelou a nova condição da língua e garantiu nova legislação educacional que permitiu o ensino bilíngue para a população de língua espanhola. Conforme Grinevald (op. cit., p. 131), “as outras línguas guaranis consistem nas variedades faladas como línguas nativas por populações indígenas, na maioria rurais, na parte oriental do Paraguai, no sudoeste do Brasil e na parte sudeste da Bolívia [...]”. No caso boliviano, há uma população aproximada de 50.000 pessoas envolvidas no esforço de décadas voltado à manutenção da variedade linguística local e para o reconhecimento oficial. O mapa apresentado a seguir representa a distribuição da população indígena da América Latina analisada anteriormente.



Figura III: As línguas mais faladas na América Latina<sup>42</sup>

Como vimos, a situação linguística na América Latina revela o convívio crescente da língua espanhola com as línguas minoritárias, com destaque para as línguas indígenas. Qualquer ação de cunho político-social que pretenda intervir nessa realidade exige conhecimento e domínio adequado de informações sobre o número de povos indígenas, o ambiente geográfico e social desses povos e o número de falantes monolíngues e bilíngues de línguas indígenas. A separação conceitual de etnia, língua e território é fundamental e deve refletir-se nas perguntas feitas em censos populacionais, pesquisas domiciliares ou qualquer outro estudo de intenção sociolinguística para uma compreensão mais satisfatória dessa realidade.

Conforme Moreno Fernández (2006), a história social das línguas indígenas hispano-americanas revela que após a redução quantitativa dos povos e dos falantes, derivada do processo de conquista e colonização, houve também um rebaixamento no uso das línguas indígenas em contexto em que se adotou o espanhol como língua oficial de suas colônias. A situação dessas línguas nos últimos cem anos tem se estabilizado em meio a um processo de lento declínio. Diante disso, pode-se dizer que a atual fase das línguas indígenas nos países sul-americanos é marginal em termos quantitativos e socioculturais. Contudo, no âmbito político nota-se uma tendência para o reconhecimento de alguns direitos indígenas a partir das reformas constitucionais

<sup>42</sup> Figura retirada e adaptada de IPOL. Disponível em: <<http://ipol.org.br/as-linguas-da-america-latina-e-sua-importancia-no-mundo-espanhol-portugues-e-linguas-indigenas>>. Acesso em: 10 out. 2020.

aprovadas nos anos 1980-90.

É perceptível que nos últimos anos as demandas dos povos indígenas têm sido verbalizadas mais vezes na língua do conquistador do que na própria língua. A diversidade linguística é um legado que a América Latina deve preservar e sua defesa deve ser compatível com o domínio da língua do conquistador para fins de comunicação e interação entre os diversos povos indígenas que vivem nesses territórios, porém isso deve ser feito de modo que a língua hegemônica não provoque o apagamento da língua nativa.

## 2.3 – Línguas indígenas no Brasil

### 2.3.1 – Resultado da conquista

Ao aportar no Brasil, no final do século XV, os conquistadores se deslumbraram com algo mais do que as belas paisagens que viram. Havia naquela grande extensão de terras uma miríade de povos que acumulavam, milenarmente, um vasto, rico e diverso conjunto cultural e linguístico. Há vários documentos que registram esse espanto, muitos dos quais escritos por missionários que seguiam os portugueses para o interior do continente em empreitadas que lhes abriam o leque do tesouro de diversos povos, costumes e línguas existentes. Esse vasto repertório de diversidade deu o fôlego necessário para que sucessivas levas de estrangeiros, durante séculos, adentrasse esse território.

Os pesquisadores, de modo geral, concordam que os indígenas encontrados pelos portugueses, no final do século XV, giravam em torno de 3 a 5 milhões de indivíduos. As densidades populacionais variavam conforme a região. O local onde os conquistadores aportaram era habitado por milhões de autóctones concentrados no litoral, vivendo em aldeias formadas por populações oscilando entre 300 e 2000 indígenas. Estes estavam divididos em dezenas de grupos tribais que somados poderiam chegar a um milhão de pessoas, número similar, na época, à população de Portugal (RIBEIRO, 1997 apud P. ABREU, 2008, p. 7).

Estudos apontam que nas várzeas dos grandes rios amazônicos e nos litorais haviam diversos grupos étnicos, com destaque para os índios Tapajós e Kambeba. Nessas áreas, “[...] relatos indicam populações ainda maiores. Conforme consignam muitos depoimentos da época, destacadamente o de Carvajal (cronista de expedição de Orellana que, em 1541, desceu o grande rio), surpreendia a considerável

densidade populacional do Amazonas, abarrotado com enormes povoados, sucedendo-se sem interrupção. Descrevendo um deles, que situou entre as vizinhanças da foz do rio Purus e 100 km acima do Negro, Carvajal afirma que teria ‘muitos bairros, cada qual com seu desembarcadouro e todos eles apinhados de índios; [e] estendia-se por mais de duas léguas e meia [com] muitos pomares e mais de quinhentas casas’<sup>43</sup>.

Linhares (1994) afirma que pesquisas arqueológicas demonstram que as várzeas amazônicas, em tempo muito recuado à chegada europeia, abrigaram denso povoamento “em estágio avançado de cultura material e organização social”, marcada por “organização tribal estável, horticultura eficiente de tubérculos tropicais e fabricação de cerâmica desde o terceiro milênio antes de Cristo que se estendeu até o início da era cristã, quando então apareceram os cacicados como forma de organização social e política”.

De acordo com Duarte (2016, p. 29), os índios tupinambás foram outro grupo que também ocupou ampla faixa litorânea do território. Pesquisas etnográficas indicam que os aldeamentos tupinambás eram formados por uma população significativamente grande na época, localizando-se em área que, atualmente, vai do estado do Pará até o Rio de Janeiro. Deve-se desconsiderar, nesse espaço, a fronteira entre o Ceará e o Maranhão, a região da foz do rio Paraíba e a região que divide o sul da Bahia do norte do Espírito Santo, haja vista que nessas regiões predominaram populações indígenas (Tapuias) etnicamente identificadas com o Tronco Macro-Jê.

As contribuições de estudos arqueológicos e antropológicos permitem que as investigações desse período remoto nos lembrem, com algum embasamento teórico e empírico, que a costa brasileira abrigou uma população indígena não homogênea em termos linguísticos e culturais. No marco da chegada dos portugueses, o povo tupinambá ocupava toda a costa, de norte a sul, em aldeias estabelecidas nas margens dos grandes rios que invadiam florestas tropicais e subtropicais. No litoral, a população chegou a, aproximadamente, 1 milhão de indígenas, número que de acordo com Fausto (2005) representa nove habitantes por quilômetro quadrado. O Brasil Central, nesse período, ocupado por povos pertencentes ao Tronco Macro-Jê, possuía aldeias com populações médias de mil habitantes, número que contrasta com a situação populacional atual, em que, em muitos casos, a densidade demográfica não

---

<sup>43</sup> Ibid., p. 8.

ultrapassa uma centena de pessoas por aldeia.

A população que era densa nas regiões citadas, começou a se desfazer a partir do momento que os portugueses intensificaram a ocupação e a exploração a leste e oeste do território com frentes de trabalho. Conseqüentemente, as populações autóctones que habitavam as regiões nordeste, sudeste e sul foram reduzidos a poucos grupos étnicos como os Xacriabá, Fulniô, Pataxós, Maxacali, Krenak, Kaingang, Xokleng, Tupiniquim e Guarani Mbyá. Além disso, a maioria dos povos que habitavam a calha do rio Amazonas e a região da ilha de Marajó não existem mais. Esses estudos explicam o porquê das regiões exploradas durante um maior espaço de tempo, atualmente, possuírem o menor número de populações indígenas, e, conseqüentemente, menos línguas originais. O fato das maiores concentrações de terras indígenas brasileiras localizarem-se no centro-oeste e na região amazônica é resultado do desaparecimento da maior parte dos nativos e das poucas terras indígenas que ainda existem a leste do território brasileiro (DUARTE, 2016, p. 30).

### 2.3.2 – Características populacionais

Segundo Melatti (2007), diversos pesquisadores fizeram estimativas sobre a população indígena no período pré-cabralino. Ele destaca três deles que escreveram suas pesquisas em períodos distintos. O primeiro é Julian Steward que publicou a obra *Handbook of South American Indians*, formada por sete volumes publicados no período de 1946-1959. Esses estudos consideraram os recursos naturais de cada região, a tecnologia usada pelas sociedades indígenas para explorá-las e o número de habitantes fornecidos pelos cronistas que as tinham visitado; com isso, calcularam uma população de 1,1 milhões de indígenas. Em obra subsequente, em co-autoria com Louis Faron, Steward lançou *Native Peoples of South America*, obra na qual foram reajustadas a densidade de algumas áreas em relação à publicação anterior, de modo que a população do Brasil passou a ser de 2.188.970 nas florestas tropicais, e de 387.440 para o Brasil Oriental.

Em outro estudo, Melatti (2004, p. 33) destaca que William Denevan, em 1976, republicou o artigo *The Original Population of Amazonia* em que reproduz o método de Steward, ajustando as densidades demográficas de algumas áreas. O total da população indígena segundo esse estudo é de 4.277.547, ou seja, quase o dobro do que foi aferido por Steward e Faron. Por fim, há o levantamento de John Hemming,

exposto no livro intitulado *Red Gold, de 1978*, no qual é apresentada estimativa do número de indígenas existente na chegada dos europeus. Para os cálculos de sua pesquisa, Hemming utiliza textos antigos que tratam dos contatos com os indígenas no período colonial até a expulsão dos jesuítas, e o total da população indígena a que chegou é de 2.431.000 indivíduos. Seguem, abaixo, os quadros demonstrativos com os resultados de cada um dos levantamentos citados.

	<i>Área km²</i>	<i>Densidade hab/km²</i>	<i>Habitantes</i>
Várzea amazônica	80012	14,6	1 168 175
Terra firme amazônica	3 920 611	0,2	784 122
Litoral (do norte até RJ)	105 000	9,5	997 500
Interior do Nordeste	477 500	0,5	238 750
Cerrado	2 178 000	0,5	1 089 000
<i>Total parcial</i>	<i>6 761 123</i>		<i>4 277 547</i>
Falta calcular	1 786 280		
<i>Superfície total do Brasil</i>	<i>8 547 403</i>		

Quadro III: População indígena pré-cabralina adaptado de Denevan<sup>44</sup>

<i>Região</i>	<i>População</i>	<i>Região</i>	<i>População</i>
Rio Grande do Sul	95 mil	Xingu, Iriti	66 mil
Santa Catarina e Paraná	152 mil	Bacia do Tapajós	86 mil
Estado de São Paulo	146 mil	MT central, vale do Guaporé	71 mil
Mato Grosso do Sul	118 mil	Bacia do Madeira	78 mil
Guanabara, Rio de Janeiro	97 mil	Amazonas, abaixo do Negro	60 mil
Espírito Santo, Ilhéus	97 mil	Amapá	49 mil
Minas Gerais	91 mil	Norte do Pará	55 mil
Bahia	149 mil	Rio Negro	98 mil
Vale do São Francisco	100 mil	Rio Branco	33 mil
Litoral do Nordeste	208 mil	Içana, Uaupés	19 mil
Interior do Nordeste	85 mil	Içá, Japurá	19 mil
Maranhão	109 mil	Solimões	47 mil
Rio Pará, Marajó	87 mil	Purus, Acre	30 mil
Tocantins, Araguaia	101 mil	Juruá, Jandiatuba, Ituí, Javari	22 mil
<i>Total do Brasil</i>			<i>2.431.000</i>

Quadro IV: População indígena pré-cabralina adaptado de Hemming<sup>45</sup>

De acordo com John Monteiro (1994) apud Melatti (2004, p. 35), a análise dessas estimativas deve considerar os equívocos de cálculos que incluem em

<sup>44</sup> Quadro retirado e adaptado de Melatti (2007, p.46).

<sup>45</sup> Quadro retirado e adaptado de Ibid., p. 47.

algumas regiões, por volta de 1500, povos indígenas que para elas se deslocaram ou nelas se constituíram somente no período colonial; bem como a transferência da densidade atual para a época que os conquistadores chegaram. Se fizermos uma comparação entre a densidade demográfica e linguística do período colonial com a atualidade, a conclusão é que vivemos catástrofes humanitárias sucessivas por cinco séculos. Dos milhões de indígenas que habitavam o Brasil entre o final do Século XV e o início do XVI restaram, em nossos dias, aproximadamente 340 mil. No mesmo contexto temporal, das 1200 línguas que existiam, sobraram somente 180, redução equivalente ao desaparecimento de 85% de nosso estoque linguístico, representando, em cálculo simples, a perda de 1000 línguas em 500 anos, ou seja, uma média de 2 línguas por ano (RODRIGUES, 2005, 1999).

Em referência ao artigo intitulado *Culturas e Línguas Indígenas do Brasil*, publicado em 1957, Melatti (2007) afirma que Darcy Ribeiro ao trabalhar com diversos temas faz um ensaio específico voltado ao cálculo da população indígena em que classificou 143 grupos que acreditava existirem, dividindo-os em "classes de grandeza". Desses cálculos, gerou-se uma faixa populacional que oscilava entre 68.100 (mínima) e 99.700 (máxima). Esses números ganharam credibilidade de modo que, em 1960, o Instituto Indigenista Interamericano (III) o adotou como a população indígena brasileira.

Nos anos subsequentes, iniciativas semelhantes tiveram andamento como a feita pelo Conselho Nacional de Proteção aos Índios (CNPI) em recenseamento que após ter sido iniciado não se completou (LAND & ALMEIDA, 1979, p. 340). Em outra tentativa, o *Summer Institute of Linguistics* (SIL), em 1967, publicou dados mais atualizados. Em simpósio ocorrido em Barbados, em 1972, em trabalho sobre os países sul-americanos não andinos, pesquisadores trataram da situação dos grupos indígenas e estimaram suas populações; as informações demográficas sobre o Brasil originavam-se de fontes dispersas e nem sempre convergentes. A década de 1980 tem o Centro Ecumênico de Documentação e Informação (CEDI) como referência na publicação de dados sobre terras e população indígena brasileira, tarefa que, posteriormente, foi herdada pelo Instituto Socioambiental (ISA), oriundo da união do CEDI com outras instituições. A forma pela qual o ISA acumula os dados populacionais segue assemelhada a utilizada por Ribeiro (1957), pois o órgão utiliza fontes para coleta de dados variadas como a Fundação Nacional do Índio (FUNAI), antropólogos, organizações não governamentais, associações indígenas, a Fundação Nacional de

Saúde (FUNASA), o Ministério da Educação e Cultura (MEC) e a colaboração dos próprios indígenas e pesquisadores localizados nas aldeias através do correio, telefone e, mais recentemente, da Internet. Por tudo isso, diversos antropólogos, linguistas, sociólogos e demais pesquisadores consideram os dados compilados pelo ISA menos incompletos, portanto, mais confiáveis.

Esse breve recorte temporal expõe o esforço da Antropologia, na segunda metade do século XX, no cálculo do número de indígenas no Brasil. Apesar de toda a precariedade dos métodos e dos resultados apresentados, a carência de uma projeção da densidade demográfica dessa população era tão grande que os dados não só foram aceitos, mas também reproduzidos por outras instituições. O que se viu nos anos seguintes foi a chegada de instituições que a partir de iniciativas particulares e metodologias científicas apropriadas desenvolveram novas pesquisas geradoras de resultados importantes que serviram como referência para outros estudos. A entrada de instituições profissionalmente robustas, entre as décadas de 1980 e 1990, como o Instituto Brasileiro de Geografia e Estatística (IBGE)<sup>46</sup>, instituição especializada em estudos estatísticos e demográficos, e o ISA, voltado para questões indígenas mais amplas, mas com um setor dedicado à distribuição geográfica dessas populações, tornaram a tarefa dos especialistas que lidam com dados demográficos um pouco mais segura.

Um mérito a ser destacado a partir da publicação do artigo de Ribeiro (1957) é o desmonte da crença de que o desaparecimento dos povos indígenas estava intimamente relacionado à mestiçagem com europeus e africanos. O fato é que as explicações para a sua devastação foram deslocadas de questões como guerras e escravidão e direcionadas para os efeitos das moléstias contagiosas para as quais esses povos não tinham defesa. Apesar de não ter sido a intenção do autor, as críticas que fizera tiveram como efeito o despertar da consciência de que os indígenas estavam inexoravelmente fadados ao contínuo desaparecimento e extinção. Contrariando esse cenário sombrio, porém, pesquisas posteriores assinalam que a densidade demográfica da população indígena passa por um processo estável de crescimento. Em torno desse pensamento está Gomes (1988) apud Coimbra Jr. (2005, p. 15), ao afirmar que: "[...] o que surge como mais surpreendente e

---

<sup>46</sup> Cf. Instituto Brasileiro de Geografia e Estatística (IBGE). Disponível em: <<https://indigenas.ibge.gov.br/estudos-especiais-3/o-brasil-indigena#>>. Acesso em: 15 set. 2020. Conforme o endereço na Internet do IBGE: “desde 1991 o Censo Demográfico coleta dados sobre a população indígena brasileira com base na categoria indígena no quesito cor ou raça [...]”

extraordinário nas relações entre os índios e o Brasil é a possível reversão histórica na demografia indígena. Certamente seria temerário afirmar que os índios, afinal, sobreviveram, e que esta é uma realidade concreta e permanente. Mas o fato é que há fortes indícios de que as populações indígenas sobreviventes vêm crescendo nas últimas três décadas, surpreendendo as expectativas alarmantes de tempos atrás". Esta ideia está em linha com Ricardo (1996, p. xii) quando este diz que "[...] foi afastada a hipótese de desaparecimento físico dos índios no Brasil, portanto, não estamos diante de uma 'causa perdida' como se chegou a dizer anos atrás". Por fim, temos a opinião de Azevedo (2000, p. 80) reforçando essa tendência, ao notar que "[...] se constata que a maioria dos povos indígenas tem crescido, em média, 3,5% ao ano, muito mais do que a média de 1,6% estimada para o período de 1996 a 2000 para a população brasileira, em geral".

Pelos debates, reflexões e dados expostos acima, podemos concluir, afirmando que já faz parte do senso comum, que as populações indígenas existentes no território nacional, de modo geral, estão crescendo, consistentemente, ao longo do tempo. Porém, é importante esclarecer que o incremento da densidade demográfica desses povos não significa que o amplo e complexo arcabouço cultural e linguístico produzido, em milênios, também esteja protegido e imune a qualquer ameaça de destruição e desaparecimento. A ideia de aumento demográfico não representa, automaticamente, a garantia que as línguas indígenas estão protegidas de ameaças ou que todos os mecanismos de sua sobrevivência sejam mantidos e utilizados pelos índios, para isso é preciso que, entre outras coisas, as línguas maternas sejam ensinadas nas escolas indígenas, que ocorra a transmissão intergeracional e que as crianças aprendam e utilizem a língua de seus antepassados.

### 2.3.3 – Línguas ameaçadas e documentação

O quadro apresentado de conscientização geral sobre a gravidade do problema relacionado à diversidade linguística deu novo sentido, reconfigurou e reestruturou o trabalho de documentação linguística que passou a ter o envolvimento de diversas instituições e estimulou a criação de novas tecnologias no campo da Informática, conseguindo aportar novos recursos para pesquisas. Em suma, o resultado é o estudo, a análise e a proposição de alternativas ao problema das línguas minoritárias ameaçadas presentes em algumas iniciativas destacadas por Austin & Sallabank (2011, p. 2), entre as quais:

- O *Hans Rausing Endangered Languages Project*, financiado pela *Arcadia*, concede bolsas de pesquisa para projetos de documentação de línguas ameaçadas, mantém um arquivo digital de gravações, transcrições e metadados e administra um programa acadêmico com mestrado e doutorado recém-introduzidos para treinar linguistas e pesquisadores.
- A *Volkswagen Foundation* financia o *DOBES Program* no sentido do desenvolvimento de projetos e pesquisas de documentação das línguas ameaçadas desde os anos 2000, juntamente com o estímulo à criação de teorias e metodologias de arquivamento em acervos digitais, estabelecendo padrões nesse seguimento.
- A NSF, a NEH e a DEL representam o esforço plurianual para preservar os registros das principais línguas antes que sejam extintas.
- O EuroBABEL, voltado para as línguas ameaçadas de extinção em que, o principal objetivo é promover pesquisas empíricas sobre línguas pouco descritas e pequeno número de falantes.
- O *Sorosoro Program of the Chirac Foundation* direcionado para o desenvolvimento sustentável e o diálogo cultural para que as línguas do mundo possam prosperar.
- O *World Oral Literature Project*, com sede na *Cambridge University*, “para registrar as vozes dos mundos desaparecidos”.

- Por fim, iniciativas menores, sem fins lucrativos, como a FEL, o ELF e a GBS.

Nesse cenário de defesa das línguas ameaçadas, desde a primeira década dos anos 1990, a documentação linguística tornou-se instrumento fundamental, concentrando-se nas atividades de campo direcionadas à gravação, processamento, arquivamento e preservação a longo prazo do corpus documental dessas línguas, com destaque para as línguas indígenas. Assim, iniciamos os trechos seguintes com um mergulho nos meandros desse conjunto de métodos que vem sendo largamente utilizado pelos pesquisadores cientes que a velocidade da tecnologia associada aos seus conhecimentos teórico-metodológicos pode ser uma eficiente aliada para um problema cuja solução não pode mais ser postergada.

## 2.4 – Propriedades da documentação linguística

### 2.4.1 – Discussão teórica

A documentação linguística é voltada à utilização de métodos, ferramentas e bases teóricas para compilar um registro multifuncional, representativo e duradouro de uma língua natural ou de uma de suas variedades na comunidade de fala. Cobre aspectos sociais locais, com foco em evidências no uso da língua, na prática social, e na faculdade cognitiva, incluindo amostras da língua oral e escrita (HIMMELMANN, 2006: v). Representa também a resposta dos linguistas à questão das línguas ameaçadas através da construção de corpus de dados primários<sup>47</sup> criados a partir da interação, cooperação e participação de pesquisadores e comunidade de fala, garantindo-se o acesso às informações a não especialistas, estudantes, acadêmicos, indígenas e público em geral por meio da Internet.

Os critérios da linguística documental permitem que os fatores determinantes para a documentação de determinada língua tenham impacto direto na consolidação e ampliação das bases empíricas de várias disciplinas (Linguística, Linguística descritiva, Tipologia linguística, Antropologia cognitiva, Sociologia, entre outras),

---

<sup>47</sup> Cf. Himmelmann (2006, p. 11): Os dados primários constituem o núcleo da documentação da língua e incluem gravações de áudio ou vídeo de um evento comunicativo (uma narrativa, uma conversa, etc.), mas também as notas feitas em uma sessão de elicitación ou uma genealogia escrita por um falante nativo alfabetizado. Esses dados primários são compilados em um corpus estruturado e devem ser disponibilizados por vários tipos de anotações e comentários.”

permitindo-lhes assim a utilização de dados de comunidades de fala pouco conhecidos. Dentre esses critérios, destacam-se: 1) o perigo de extinção da(s) língua(s); 2) o uso racional e maximizado dos recursos destinados à pesquisa; 3) e a possibilidade de verificabilidade e análise dos dados. Ressalte-se a importância do primeiro item como o princípio básico norteador das preocupações, pesquisas e projetos ancorados nos fundamentos da documentação linguística.

Assim, a formação dos acervos digitais através da documentação exaustiva deve se contrapor a possibilidade de criação de “cemitérios” ou “armazéns de dados” inúteis, ou seja, deve transformá-los, principalmente, em instrumentos de conservação das línguas ameaçadas ao apoiar a pesquisa interdisciplinar, de baixo custo e acessível, formada por base de dados aberta a pesquisadores de modo a facilitar sua difusão. Além dessas observações, Himmelmann (2007) alerta que esses projetos não devem prescindir da característica basilar de uma documentação linguística duradoura e multifuncional, a criação de coleções documentais cujos acervos abertos sejam conformados, por etapas, consoante a sucessiva compilação e integração de conjunto de dados primários verificáveis, oriundos de vários projetos individuais diferentes ao longo do tempo. Junto à preocupação com a integração dos dados primários dos pesquisadores para a formação de grandes acervos, destacam-se também outros fatores nesse processo documental (vistos, detalhadamente, mais adiante), como sua coleta, verificabilidade, armazenamento e preservação; bem como a formação de equipes multidisciplinares e a cooperação entre pesquisadores e comunidade.

Além da definição tradicional acima, há alguns aspectos da documentação linguística que contribuem para uma melhor compreensão de suas finalidades e princípios. A seguir destacamos quatro deles: 1) trata-se de método voltado ao registro amplo das características das práticas linguísticas de determinada comunidade de fala. Exigindo para isso, a aplicação de técnicas analíticas para a transcrição, tradução, anotação e publicação desse registro. É um erro, portanto, imaginar a DL como uma simples coleta de material para preservação linguística sem qualquer tipo de esforço analítico (HIMMELMANN, 1998, p. 2). Muitos ainda confundem a linguística descritiva<sup>48</sup> com a documentação linguística ao imaginar que esta emula as práticas

---

<sup>48</sup> A Linguística descritiva ou descrição linguística refere-se ao trabalho de análise e descrição objetiva das formas linguísticas usadas por determinada comunidade de fala. Modernamente, baseia-se na abordagem estrutural da linguagem. O dicionário Routledge a define de duas formas: “1. No seu sentido mais restrito, um termo para as abordagens inseridas no estruturalismo americano representado por L.Bloomfield, Z.S.Harris, H.A. Gleason, e outros, onde o rótulo "descritivo" acentua

da primeira, consagradas desde o início do século XX, incorporando, simplesmente, novidades computacionais no processo de documentação. Tais considerações ignoram a teoria e metodologia própria, assim como a produção e importância do corpus de dados e metadados nas análises (LÜPKE, 2010); 3) outra confusão marcante se refere ao seu escopo, alguns limitam-se a imaginar que sua utilização e abordagem restringe-se somente às línguas ameaçadas, ignorando que seus princípios e práticas podem ser aplicadas a todas as línguas, grandes ou pequenas, em risco de extinção ou não (AUSTIN, 2010); 4) por fim, a interpretação de que a DL, como se apresenta, é algo passageiro, uma distorção da Linguística que logo será superada. Esse tipo de interpretação negligencia, como veremos, na verdade, tratar-se de uma mudança de paradigma que veio para ficar.

Conceitualmente cientes do que é a documentação linguística, ressaltamos que seus princípios diferem da linguística descritiva, considerando que esta prioriza o registro de uma língua a partir de um sistema de elementos abstratos e regras no qual o resultado é a produção de gramáticas, dicionários e anotações. Por sua vez, a documentação linguística é, por natureza, multidisciplinar e, como observa Woodbury (2011), não se restringe à teoria e aos métodos da linguística descritiva, mas se baseia em “conceitos e técnicas da Linguística, Etnografia, Psicologia, Ciência da computação, Artes Cênicas e muito mais”.

Há vários debates entre os acadêmicos sobre a necessidade de uma teoria geral sobre a documentação linguística. Grenoble & Furbee (2010), por exemplo, afirmam ser prematuro dizer que a documentação linguística possui uma "teoria". Segundo ela, existe, na verdade, a ênfase em melhores práticas e na ética, principalmente no trabalho de campo. No que se refere à questão da compilação de dados primários, apesar de Himmelmann (2006, p. 1-4) afirmar que a tarefa para ser executada não precisa estar obrigatoriamente associada a qualquer projeto teórico ou prático, sua ideia não é desprezar a importância de ambos, mas mostrar que a teoria

---

vários aspectos: (a) linguística sincrônica (sincronia vs diacronia) no sentido de Saussure (1916), ou seja, sem referência a contextos históricos; (b) descrição de línguas individuais através da generalização da análise de corpus (por exemplo, procedimentos de F. Boas na investigação de línguas indígenas americanas), por oposição à construção de gramáticas universais; (c) procedimentos empíricos, positivistas (empirismo), ou seja, inventário objetivo baseado na observação com análise distributiva (distribucionalismo). 2. Também gramática descritiva. No seu sentido mais amplo, qualquer tipo de descrição não prescritiva ou não normativa de diferentes variedades linguísticas, que codifica as regularidades conforme o uso (BUSSMANN et al., 1998, p. 296).

está presente nos projetos multidisciplinares que aplicam os métodos essenciais para o registro, processamento e preservação de dados linguísticos primários, de modo que a coleta seja realmente útil para propostas teóricas de diferentes finalidades, realizadas por diferentes disciplinas.

Entretanto, é importante salientar que foram apontados graves problemas teóricos na documentação linguística, entre os quais destaca-se a falta de um limite temporal para o registro de eventos comunicacionais na comunidade de fala. Enquanto a prática prega que se registre o máximo possível desses eventos sem qualquer restrição de tempo; na teoria, os aspectos balizadores são determinados ou não pelos objetivos do projeto. A escolha de um “objetivo indeterminado” forma uma abundância de exemplos para cada tipo de evento comunicativo que exige a descoberta da tipologia desses eventos em cada comunidade de fala. Essa dificuldade é tratada pela sociolinguística por meio da etnografia da comunicação. Diante disso, na falta de teoria que norteie o registro das diferentes representações de categorias e concepções das diversas línguas, a sugestão é que a documentação priorize, além da coleta de dados, a criação de interfaces e variados tipos de análises (HIMMELMANN, 2007).

Tornar essas interfaces transparentes e teoricamente independentes facilita o arquivamento da corporal documental em formatos relacionáveis, fornecendo um conjunto de dados comparativos preparados de maneira semelhante em grande escala (HIMMELMANN, 2007, 2006; WHALEN, 2004). A partir dessa breve apresentação das questões teóricas candentes da documentação linguística, vejamos como seus fundamentos assentados na abordagem teórica, metodológica e prática se configuram.

Visto como um dos principais teóricos e pensadores acerca dos fundamentos da documentação linguística, Himmelmann (2007) destaca dois métodos essenciais para o registro de dados primários: O comportamento linguístico observável e o conhecimento metalinguístico. Vejamos como funcionam.

O comportamento linguístico observável, relacionado à linguística documental, refere-se ao registro duradouro e multifuncional de amplo conjunto de dados primários que representam a língua utilizada em dado momento, em determinado local, por uma comunidade que a expõe nos mais diferentes sentidos e possibilidades. Himmelmann (2007, p. 22) destaca a importância desses registros conterem amplas amostras desse comportamento que explicitem o modo pelo qual as pessoas se comunicam

realisticamente em seu cotidiano, priorizando os diálogos que acontecem na comunidade de fala, desde as conversas banais e rotineiras, os complexos rituais, os diálogos de pais e filhos, até as discussões políticas entre os anciões. Obviamente esses registros devem ser acompanhados dos filtros teóricos e éticos. A documentação do comportamento linguístico observável não é o único método capaz do registro duradouro e multifuncional da língua, por haver aspectos desta, não acessíveis e perceptíveis. Nesse caso, deve-se aplicar o conhecimento metalinguístico utilizado, principalmente, pela linguística descritiva.

O conhecimento metalinguístico pode ser definido como a capacidade desenvolvida pelo falante nativo de expressar sua compreensão das particularidades do comportamento linguístico de seu povo. Enfim, seria a capacidade de alguns falantes interpretarem e retratarem seus eventos comunicacionais cotidianos. O processo de documentação do conhecimento metalinguístico caracteriza-se: (i) pela existência de uma taxonomia de base linguística (sistemas de parentesco) e outra de base popular (plantas, animais, instrumentos musicais, expressão de número, paradigmas morfológicos); (ii) pela ausência de problemas de ordem teórica e ética graves; (iii) e por um processo de documentação que por vezes torna-se difícil pela dificuldade de acesso ao falante. O conteúdo do corpus de dados primários explicita o conhecimento metalinguístico dos falantes nativos cujas práticas linguísticas refletem “situações registráveis”, conforme o método escolhido. A documentação resultante constitui-se de informação básica e suficiente para a escrita de gramáticas descritivas e dicionários (HIMMELMANN, 2007).

Como vimos, a documentação do conhecimento linguístico observável e a documentação do conhecimento metalinguístico são métodos utilizados para a obtenção dos registros que constituirão o corpus de dados primários relativos à documentação dos experimentos interativos linguísticos de falantes nativos.

Até o final da primeira década dos anos 2000, havia poucas experiências práticas confiáveis sobre a estruturação e manutenção desse tipo de corpora digital. Uma exceção era o modelo proposto pelo Programa DOBES, - o qual, atualmente, podemos dizer é um paradigma - cuja estruturação dos dados primários é feita, basicamente, de duas maneiras: 1) registrando individualmente os eventos comunicacionais, denominados “sessões” (chamados também de “documentos”, “textos” ou “grupos de recursos”); e 2) criando bases de dados léxicos, conforme três processos: compilação de dados; transcrição de gravação; e anotações do campo.

Himmelman (2007, p. 26) defende a utilização de base de dados léxicos como formato para organizar dados primários porque a reunião de toda a informação disponível para cada entrada de léxico garante que o significado e as propriedades formais dessa informação sejam compatíveis e, principalmente, pelo fato de a lista de léxicos ser um recurso extremamente útil quando se trabalha com transcrição e tradução de gravações. Essas tarefas ganharam nova dinâmica com o surgimento de softwares voltados à automatização dessas atividades. Teremos uma seção específica para tratar das ferramentas de TI essenciais para quem trabalha com documentação linguística.

Um projeto de documentação completo é constituído por amplo conjunto de dados primários que deve ser acompanhado por observações linguísticas e metalinguísticas. Porém, esse grande volume de dados primários tem utilidade somente quando é acessível àqueles que não tiveram envolvimento com sua compilação e quando possui informações de tipo variado que auxiliem o pesquisador. Nesse caso, os tipos de informações primordiais, entre outras, acerca do conteúdo de um corpus de dados primários são três: metadados, sessões e anotações. Vejamos cada um deles:

1) os metadados que possuem informações sobre o projeto de documentação, equipe, objetivo (nome da variedade da língua, onde é falada, número e tipo de registos, etc.). Essa categoria de metadados possui duas funções que agem de forma diferente, mas se complementam: catalogadora e organizativa. A primeira funciona de forma análoga ao catálogo de uma biblioteca, por atuar administrativamente junto à estrutura do corpus digital, auxiliando os procedimentos informáticos. A função organizativa define a estrutura do corpus e permite sofisticadas ações, como a execução de diferentes tipos de buscas por filtro no interior desse corpus ou de vários se estiverem no formato digital. As extensões digitais relacionados aos metadados citados são: o *Open Language Archives Community* (OLAC) cuja função catalogadora acelera o acesso à maioria dos repositórios de dados primários existentes, digitais ou não; e o padrão *International Standard for Language Engineering* (ISLE) ou *ISLE Meta Data Initiative* (IMDI) que incorpora todas as funcionalidades do padrão OLAC, sendo compatível como ferramenta de administração de corpus utilizada, principalmente, na documentação de línguas digitalmente arquivadas (NATHAN, 2010; AUSTIN, 2007).

2) A sessão (metadado) trata dos segmentos dos dados primários que devem ser acompanhados das seguintes informações básicas: nome único por sessão; nome

de quem gravou e de quem estava presente durante a gravação; nome do autor da gravação e do equipamento utilizado; informação sobre qualidade dos dados; nome daqueles que têm acesso aos dados da sessão; breve descrição do conteúdo da sessão; link dos diferentes arquivos que em conjunto formam a sessão (HIMMELMANN, 2007, p. 27).

Além dos metadados mencionados, há a 3) anotação<sup>49</sup>, uma sessão individual para gravação (áudio e vídeo de eventos comunicacionais) e documentação que precisa de informação adicional para tornar o corpus de dados primários acessíveis aos usuários que não conhecem a língua documentada. Não há consenso sobre nomenclatura e formato de anotações incluídas em cada sessão. Existem dois tipos de anotação, a mínima e a completa. Enquanto a mínima, mais utilizada, é composta por uma transcrição e uma tradução que acompanha todos os segmentos do corpus de dados primários ou pelo menos parte importante deles; a completa, possui (i) vários níveis de glosa interlinear, (ii) comentários gramaticais, etnográficos e referenciais cruzados, os mais exaustivos possíveis, entre as diferentes sessões, (iii) e recursos compilados em determinada documentação<sup>50</sup>.

Como vimos, a documentação linguística concentra toda a informação pertinente aos dados primários em metadados (sessão, anotação, etc.). Estes são também denominados “recursos de acesso geral” e devem possuir estrutura que contemple os seguintes pontos: 1) uma introdução geral que ofereça informação sobre os antecedentes da comunidade de fala e da língua: nome da(s) língua(s), filiação, principais variedades, etc., o entorno do trabalho de campo, os métodos utilizados para registrar os dados primários, um resumo dos conteúdos, a estrutura e o alcance do corpus de dados primários e sua qualidade; 2) esboços breves das principais características etnográficas e gramaticais que estão sendo documentadas; 3) uma explicação das diferentes convenções que se utilizam (ortografia, abreviaturas da glosa, outras abreviaturas), índices para as línguas/variedades, conceitos analíticos chave, etc.; 4) links e referências a outras fontes (livros e artigos publicados previamente sobre a variedade linguística ou comunidade documentadas, outros projetos relacionados com a comunidade ou seus vizinhos, etc.).

---

<sup>49</sup> Conforme Himmelmann (2007, p. 28), apesar da “anotação” ou “comentário” também ser um metadado, essa denominação deve ser reservada para os dados com função catalogadora ou organizativa, enquanto os outros tipos de informação que acompanham seguimentos de dados primários devem chamados de anotação.

<sup>50</sup> Ibid., p. 28.

Os princípios teóricos apresentados representam o padrão básico de documentação linguística; porém existem outros padrões de documentação oriundos da linguística estrutural, como o gramático-dicionário, cuja incorporação de alguns de seus fundamentos permitiu o surgimento de um terceiro, o padrão ampliado de documentação linguística<sup>51</sup>. Vejamos suas principais características.

Este padrão focaliza o sistema da língua, interpretando-a como um sistema abstrato de regras e oposições implícito no comportamento linguístico observável, de modo que a documentação seja representada pela compilação de uma gramática<sup>52</sup> e um dicionário<sup>53</sup>, reunidos em coleção de textos que exprimem o funcionamento do sistema. Com frequência esses textos convertem-se no corpus de dados primários baseados na descrição do sistema. Contudo, o acesso aos dados primários não é permitido. Entre as principais críticas a esse padrão estão: (i) indisponibilizar o acesso direto ao corpus de dados primários; (ii) não apresentar registros duradouros e multifuncionais da língua; (iii) não manipular corretamente os dados primários; (iv) gerar descrições reducionistas em relação aos dados primários, limitando a documentação da língua a um sistema de regras e oposições, etc. Apesar do conjunto de críticas, são reconhecidas a importância e relevância da gramática descritiva e dos dicionários em relação a seu objetivo principal, qual seja, oferecer descrição e documentação de um sistema. Como a admissão de sua funcionalidade, tornou-se possível combinar os dois padrões apresentados, gerando um terceiro<sup>54</sup>.

Esse terceiro padrão, denominado padrão ampliado para documentação linguística, como vimos, é resultado da conclusão de que a abordagem estruturalista da língua, como um sistema de regras e oposições, é compatível com a proposta de documentação em que gramáticas descritivas e dicionários oferecem representações apropriadas do sistema. Assim, é possível acreditar que um processo realmente exaustivo de documentação linguística consiga reunir não só um amplo corpus de dados primários anotados, mas também uma gramática descritiva e dicionário completos. Essa simbiose exigiu a adaptação das características próprias dos padrões básico e padrão ampliado de documentação linguística, entre as quais as que podemos ver nos quadros V e VI, abaixo, onde modificações como a inserção da

---

<sup>51</sup> Ibid., p. 29.

<sup>52</sup> Conjunto de regras para produção de enunciados.

<sup>53</sup> Listas de pares convencionais forma-significado utilizados na produção de enunciados.

<sup>54</sup> Ibid., p. 29.

anotação “análise descritiva” exaustiva e o ajuste em certos metadados (Quadro VI) determinaram a alteração da categoria “básico” para “ampliado”. A associação entre esses padrões de documentação provocou dificuldades de ordem teórica e financeira nos projetos de documentação linguística ampliada<sup>55</sup>.

Em geral, os questionamentos se concentraram, em termos teóricos e práticos, na possibilidade de se incorporar corretamente a gramática descritiva como parte integrante essencial da documentação linguística e todos os agravantes oriundos desse procedimento. Os problemas financeiros seriam uma decorrência da solução escolhida para questão teórica e prática relacionada, principalmente, à elevação dos custos e a intrínseca viabilidade de execução do projeto. Os debates levantados aqui, sucintamente, refletem a conclusão de que a documentação linguística exaustiva para ser completa precisava incorporar um elemento externo, a capacidade analítica trazida da linguística estrutural. Apesar de ainda existirem dúvidas se a análise, após a compilação dos dados, deve ser apresentada no formato de uma gramática descritiva ou inserida na documentação mediante anotações (distribuídas no corpus). O que parece prevalecer atualmente entre os pesquisadores é a consciência de que o tempo, geralmente longo, a ser empregado no campo deve ser produtivo, privilegiando-se a documentação analítica em detrimento da “pura” ou, como parece ser uma tendência, conforme o escopo do projeto, promover a combinação de padrões (HIMMELMANN 2007, 2006, 1998).

---

<sup>55</sup> Ibid., p. 30.

Dados primários	Instrumento	
Gravações / registros de comportamento linguístico observável e conhecimento metalinguístico (possíveis formatos básicos: sessão e base de dados léxicos)	Por sessão	Para toda a documentação
	<b>Metadados</b> . Lugar e data da gravação; . participantes responsáveis pela gravação; . equipe de gravação; . descritores de conteúdo <b>Anotações</b> . transcrição; . tradução; . outras glosas e comentários linguísticos e etnográficos	<b>Metadados</b> . localização da comunidade; . pessoal que trabalha no projeto de documentação; . colaboradores na comunidade; . agradecimentos <b>Recursos de acesso geral</b> . introdução; . convenções ortográficas; . esboço etnográfico; . esboço gramatical; . convenções de glosa; . índices; . links para outros recursos

Quadro V: Padrão básico de documentação linguística<sup>56</sup>

Dados primários	Instrumento	
Gravações / registros de comportamento linguístico observável e conhecimento metalinguístico	Por sessão	Para toda a documentação
	<b>Metadados</b> <b>Anotações</b> . transcrição; . tradução; . outras glosas e comentários linguísticos e etnográficos	<b>Metadados</b> <b>Recursos de acesso geral</b> . introdução; . convenções ortográficas; . esboço etnográfico; . esboço gramatical; . convenções de glosa; . índices; . links para outros recursos
		Análise descritiva . etnografia . gramática descritiva . dicionário

Quadro VI: Padrão ampliado de documentação linguística<sup>57</sup>

Neste ponto do texto já sabemos que a documentação linguística não é totalmente inovadora. A compilação de documentos e a valorização das atividades de fala já eram objeto das pesquisas de filólogos e antropólogos na passagem dos séculos XIX-XX. Porém, deve-se ressaltar que as principais novidades trazidas não fazem parte do que existe na corrente linguística dominante e muito menos existiu em

<sup>56</sup> Quadro retirado e adaptado de Himmelmann (2007, p. 30).

<sup>57</sup> Quadro retirado e adaptado de Himmelmann (2007, p. 38).

séculos anteriores. A documentação linguística vem se consolidando não apenas pelo fator primordial de registrar as línguas ameaçadas, mas também pelo desenvolvimento das novas tecnologias da informação que têm contribuído para a coleção, análise, arquivamento e compartilhamento de registros documentais de uma forma jamais vista anteriormente (NATHAN, 2010; AUSTIN, 2010; HIMMELMANN, 2007). Contudo, isso não significa que problemas inexistem. Certamente existem, como os teóricos apontados, mas também há os de tipo metodológico, prático e ético, etc. Contudo, é importante ressaltar que os avanços científicos são dialéticos, portanto, as críticas e o revisionismo fazem parte de seu processo de aprimoramento. Abaixo são representados os fluxos de trabalho conforme os princípios da linguística descritiva e da documentação linguística.

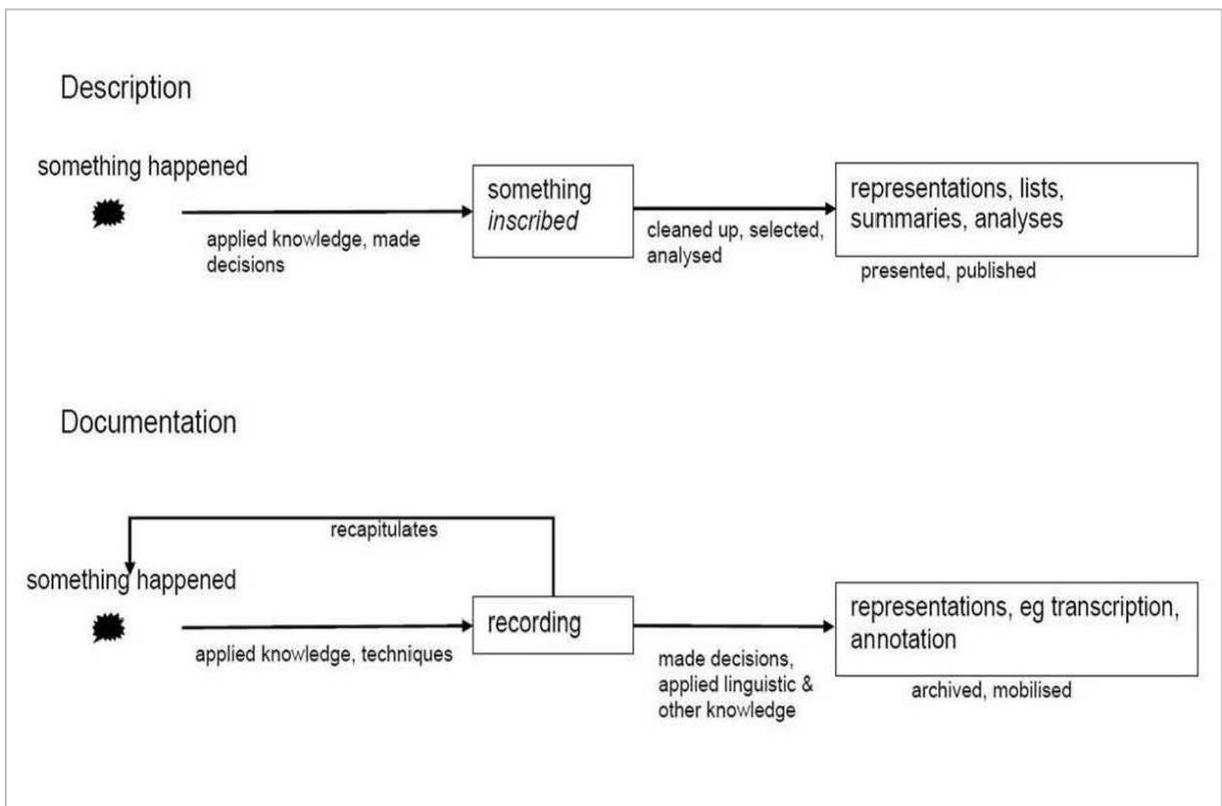


Figura IV: Fluxo de trabalho na descrição e documentação linguística<sup>58</sup>

<sup>58</sup> Figura retirada e adaptada de Austin (2010, p. 23).

Para finalizar esta subseção, ainda no campo exclusivamente da documentação linguística, em linha com as proposições de criação de potenciais coleções de corporas multifuncionais elaboradas a partir dos registros e documentação das práticas linguísticas, com a possibilidade de acesso aberto a pesquisadores e demais interessados como proposto por Himmelmann (2007, 2006, 1998), recomendamos um conjunto bibliográfico que orienta na formação de corporas de determinados tipos (narrativas míticas, genealogias familiares, folclore, culinária, etc.), 2) elicitación e 3) técnicas de captura e processamento de dados. Veja a lista abaixo onde se destacam:

- 1) As discussões existentes no livro de Grenoble & Furbee (2012) sobre estudos de caso que envolvem a criação de corpora de tipos específicos; nesta obra, no artigo “Documenting different genres of oral narrative in Cora (Uto-Aztecan)”, Soto aborda a prática da documentação linguística sobre as diferentes narrativas orais e diálogos coletados na comunidade. Há também o trabalho de Haig et al. (2011) em capítulo intitulado “Comparing corpora from endangered language projects: Explorations in language typology based on original texts” que trata da pesquisa comparativa dos variados discursos naturais falados, extraídos de arquivos digitais, a partir de uma variedade de línguas tipologicamente diferentes. Finalizando este tópico, deve-se destacar o artigo de Mosel (2014), “Putting oral narratives into writing – experiences from a language documentation project in Bouganville, Papua New Guinea”, no qual se orienta o pesquisador, a partir dos fundamentos da documentação linguística, a elaborar versões editadas das narrativas orais de povos indígenas (lendas, autobiografias, histórias, etc.).
- 2) As diversas instruções para elicitación, entre as quais sugerimos alguns trabalhos. O artigo de Mosel (2006), “Fieldwork and community language work”, com a abordagem de diversos assuntos (gênero de fala, conteúdo das gravações, criação de corpus, transcrição, tradução, etc.), com destaque para a questão das variadas práticas de elicitación empregadas na pesquisa de campo objetivando documentar o conhecimento metalinguístico. Há também o artigo de Himmelmann (1998), “Documentary and descriptive linguistics”, em obra clássica sobre diversos temas, este possui indicações de técnicas de elicitación. Encerrando este tópico, há o estudo de Bower (2015), “Linguistic Fieldwork: A Practical Guide” onde dois capítulos são reservados para instruções e análises da aplicação de variados métodos para a coleta de dados, com destaque para a elicitación do conhecimento metalinguístico.

- 3) As várias técnicas de captura e processamento de áudio e vídeo para a formação de corpora multifuncional, entre as quais sugerimos a consulta a algumas obras importantes. O artigo de Austin (2007), “Los datos y la documentación lingüística” indica o processo de formação de um corpus documental completo. No texto o autor descreve os principais processos na coleta e representação dos dados linguísticos num projeto de documentação, além das ferramentas de gravação, as tecnologias a serem empregadas e as etapas do trabalho. O trabalho de Bowerman (2015) trata desses aspectos da documentação a partir de dois capítulos (2 e 4) que abrangem métodos de captura e processamento de áudio e vídeo conforme os parâmetros de formação de corpus exaustivo e multifuncional, abrangendo o comportamento linguístico observável e o conhecimento metalinguístico. No artigo de Margetts & Margetts (2014) abordam-se questões práticas de gravações de áudio e vídeo na pesquisa de campo a partir das experiências pessoais dos próprios escritores e de outros pesquisadores. Nesse sentido, são destacados o perfil dos equipamentos a serem utilizados, a forma de utilizá-los e como explorar o máximo de seus recursos. Indicam-se também algumas instruções sobre o que registrar, a descrição do fluxo de trabalho do processamento de dados e o modo de uso dos equipamentos eletrônicos nessa operação. Para finalizar, o trabalho ressalta a importância dos equipamentos auxiliares, como os voltados ao fornecimento de energia e as mídias que armazenarão as gravações de áudio e vídeo, entre outras informações técnicas essenciais.

## 2.4.2 – Discussão metodológica

### 2.4.2.1 – Trabalho de campo

Nos tópicos seguintes veremos algumas reflexões referentes a princípios básicos de ética e direitos aplicados ao trabalho de campo na documentação linguística, principalmente aqueles voltados às línguas ameaçadas de extinção.

Esta seção avalia as principais questões relacionadas à ética e direitos ligados à pesquisa de campo, tendo em vista que o método, no planejamento, é um elemento essencial para o sucesso de uma empreitada de documentação linguística, por envolver acadêmicos, equipes de apoio e membros da comunidade de fala em ambiente de trabalho idealmente colaborativo. Essa organização é fundamental para que o linguista se concentre plenamente em todas as etapas do processo de documentação e nos objetivos finais (RICE, 2006, p. 9).

A documentação linguística é complexa e implica reunir grupos sociais distintos em empreitada que envolve objetivos comuns. A atividade de campo, muitas vezes, torna o titular da pesquisa um mediador natural de conflitos entre os participantes nas etapas que a compõem. Diante disso, princípios éticos internalizados e compartilhados pelo grupo têm um significado especial. Frente aos diferentes desafios que os projetos têm em termos de objetivos a atingir, formação de equipes, localização e relação com a comunidade de fala, existiria um padrão normativo único para as questões éticas no trabalho de campo? Esses padrões éticos comportamentais seriam plenamente executáveis empiricamente ou engessariam os objetivos do projeto, com repercussões nas especificidades do trabalho do pesquisador, com reflexos em sua investigação?

### 2.4.2.2 – Ética

A metodologia de campo evoluiu significativamente de um modelo não cooperativo, em termos de pesquisa comunitária, para um modelo cooperativo que tem, entre seus objetivos, a capacitação das comunidades de fala (CAMERON et al., 1992 apud DWYER, 2006).

Esses avanços, porém, não estão claros em relação ao padrão ético a adotar no trabalho de campo. Existem duas propostas usadas largamente e, por vezes, apresentadas como excludentes: a ética geral e a ética normativa. Serão mesmo

excludentes ou haveria a possibilidade de combinar alguns de seus princípios? A primeira, conforme Dwyer (2006), impõe ao pesquisador conhecimentos empíricos amplos, como os saberes técnico-jurídico das legislações sobre como proceder para ingressar nas áreas de pesquisa (internacionais, nacionais ou locais)<sup>59</sup>; os tipos de consentimento; as relações de reciprocidade, a necessidade de discutir o que será registrado, além da propriedade dos dados; e a atenção ao respeito às diferenças culturais de pesquisadores e falantes. Como não há uma lista de regras padronizadas, a sensibilidade do pesquisador é que deve orientar a adoção dos princípios gerais adequados, de modo a explicitá-los nas situações em que seja necessário mediar conflitos que envolvam todos os grupos sociais juntamente com as instituições acadêmicas e de fomento envolvidas no projeto. Certamente os envolvidos em projeto de documentação linguística, em algum momento, enfrentarão dificuldades, pois não há uma panaceia para todos os meandros dessa atividade (GLASER & HARDEN, 1995 apud AUSTIN, 2010). O padrão de comportamento ético a ser utilizado na pesquisa de campo, apesar da aparência genérica, serve, no primeiro momento, para indicar os princípios de mediação a serem aplicados na relação que envolve comunidades autóctones e pesquisadores de realidades culturais antitéticas.

Nos termos da ética normativa, empregados por universidades e organizações de pesquisa em geral, as decisões tomadas durante o trabalho de campo estão no âmbito da ética profissional, onde práticas normativas determinam os padrões de melhores práticas. Esses padrões solapam códigos de conduta empiricamente elaborados e podem ter implicações sérias se o uso em excesso da normatividade não considerar as fragilidades socioculturais das comunidades (DWYER 2006). A complexidade dos eventos no campo exige que códigos de conduta personalistas sejam substituídos por diretrizes e normas metodológicas permanentemente empregadas que padronizem relações de trabalho, produzam resultados de pesquisa satisfatórios e melhore a comunicação entre os participantes. Essa questão ganhou tal dimensão que “[...] cada vez mais organizações locais, como centros culturais ou

---

<sup>59</sup> No caso do Brasil, especificamente, o ingresso em Terras Indígenas (TI) é regulamentado por normativas da Funai e as autorizações de ingresso nas TI “são de competência exclusiva da Presidência da Funai, após a devida instrução do processo administrativo nos termos das referidas normativas, observando-se a anuência prévia dos representantes dos povos indígenas envolvidos, conforme dispõe os artigos 6º e 7º, da Convenção 169 da OIT bem como a manifestação das unidades regionais da FUNAI, das coordenações gerais e, quando necessário, a análise jurídica pela Procuradoria Federal Especializada – PFE/AGU”. Disponível em: <<https://www.gov.br/funai/pt-br/atuacao/terras-indigenas/ingresso-em-terra-indigena>>. Acesso em: 20 mar. 2023.

organizações não governamentais (ONGs) indígenas desenvolveram declarações ou regras éticas específicas relacionadas à condução de pesquisas [...]” (AUSTIN, 2010, p. 36). Até nas abordagens pragmáticas é fundamental que as equipes tornem transparentes suas normas éticas de pesquisa. Assim, as equipes, individualmente, devem criar códigos éticos próprios para cada área, voltados para um determinado projeto de pesquisa que inclua diretrizes sobre consulta e negociação entre a comunidade de fala e pesquisadores que envolva todas as etapas da pesquisa, desde o planejamento até a publicação.

Conforme Dwyer (2006, p. 34), em geral, princípios éticos englobam as seguintes fases da pesquisa: 1) o planejamento ético que deve ser feito para cada fase da pesquisa e engloba a apuração de desempenho das equipes, dos benefícios e possíveis prejuízos da pesquisa e da atuação dos participantes locais em todas as fases (deve-se identificar todos os participantes, inclusive instituições patrocinadoras e estimar a remuneração dos participantes locais); 2) o trabalho de campo, no qual os pesquisadores estabelecem relacionamentos, negociam acordos para obter dados e obtêm o consentimento informado; 3) a análise que trata de decisões éticas normativas como a quantidade de níveis mínimos adequados para a anotação. Essa normatividade é decorrente da anotação determinar o acesso dos materiais a públicos específicos; 4) o arquivamento, no qual o pesquisador controla a visibilidade do(s) consultor(es) e do(s) falante(s) na pesquisa (anônimo ou identificada), o acesso dos usuários aos materiais (pesquisadores, comunidade de fala e público em geral), e os materiais a serem acessados.

A utilização desses princípios éticos, na prática, exige a especificação das relações éticas e legais, entre todos os participantes do processo de documentação, diferenciadamente, principalmente no papel referente ao consultor e ao pesquisador. O consultor (falante), membro de uma comunidade linguística e administrativa onde se encontra completamente integrado, tem como função estabelecer os meios estáveis de acesso do pesquisador ao campo de pesquisa. O pesquisador, por outro lado, é geralmente classificado de acordo com suas funções na pesquisa (professor local ou estrangeiro) e com o modo pelo qual é visto pela comunidade: insider (aceito como membro da comunidade) ou outsider (de fora da comunidade). Na etapa de planejamento da pesquisa, considerar o vínculo do pesquisador com a comunidade é importante por três motivos: 1) a participação de um pesquisador local facilita a solução de problemas relacionados aos desdobramentos da pesquisa; 2) esse

pesquisador pode criar produtos (dicionários, livros, gramáticas, etc.) úteis para a comunidade e doá-los; 3) O pesquisador estrangeiro, por outro lado, pressionado pelas instituições com as quais tem vínculo (universidade e/ou agência de financiamento) pode limitar-se a colher dados linguísticos e efetuar publicações sem considerar atitudes de reciprocidade com a comunidade de fala. Em sintonia com essas reflexões, Dwyer (2006, p. 37) afirma que: “o fato desses atores – pesquisadores de campo individuais, comunidades, consórcios de pesquisa, agências de financiamento, arquivos e usuários - estarem localizados em diferentes países tem implicações legais para o armazenamento, propriedade, transferência e publicação de dados [...]. Porém, o mais importante para o sucesso ou fracasso de uma dada colaboração de pesquisa são as redes mutáveis e altamente contextuais de poder e pertencimento (insider / outsider) entre esses atores. Um projeto de pesquisa em qualquer escala faria bem em avaliar essas relações jurídicas e sociais no estágio de planejamento [...]”. A figura abaixo esquematiza o modo pelo qual essas relações ocorrem.

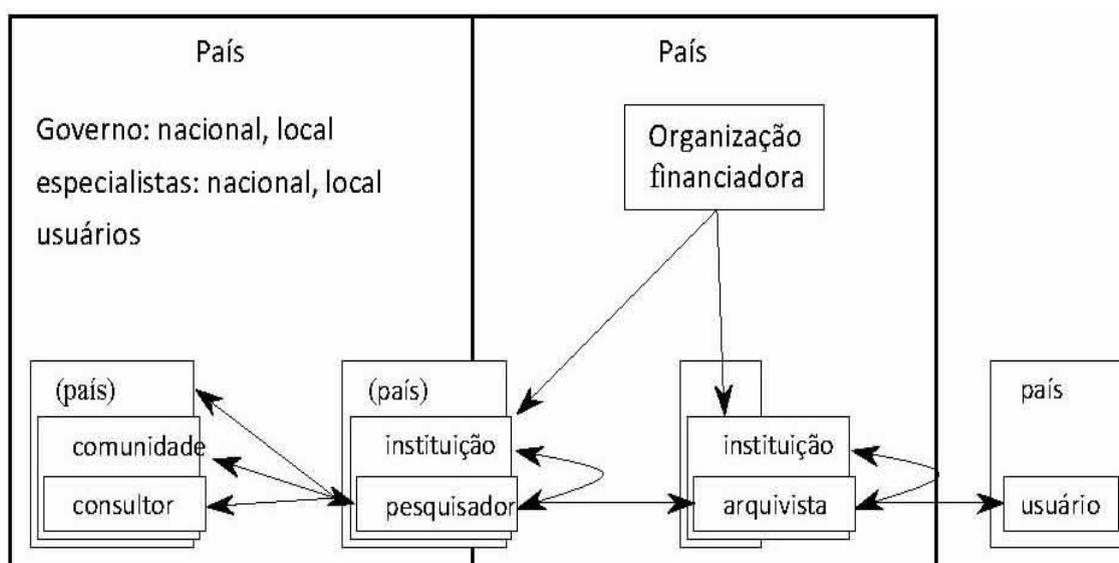


Figura V: Fluxo de participantes em trabalho de campo linguístico<sup>60</sup>

Associado ao que já tratamos sobre questões éticas na pesquisa de campo, listamos abaixo um esboço de cinco princípios éticos fundamentais: 1) não cause danos (inclusive o intencional) com as informações; 2) exercite a reciprocidade e equidade na relação entre os participantes da pesquisa em termos continuamente

<sup>60</sup> Figura retirada e adaptada de Dwyer (2006), Hiß (2001), Wittenburg (2001–2000).

consensuais e respeitosos; 3) faça algum bem (tanto à comunidade de fala quanto à ciência); 4) obtenha consentimento informado antes de começar a pesquisa; 5) armazene e difunda seus dados e resultados. Para mais informações sobre modelos de melhores práticas de pesquisa, organizações como a *E-MELD School of Best Practice*<sup>61</sup>, o *DOBES Programme*<sup>62</sup> e o *AIATSIS Research ethics committee*<sup>63</sup> fornecem guias detalhados. Após esse breve levantamento sobre os principais aspectos éticos da pesquisa de campo, é crucial refletir sobre os direitos dos participantes nessa atividade (AUSTIN, 2010; DWYER, 2006).

#### 2.4.2.3 – Direitos

Os programas de documentação linguística analisados, localizados no continente europeu (DOBES) e americano (DMA e AILLA) juntamente com suas equipes de pesquisadores estão sujeitos a pelos menos três domínios jurídicos diferentes: 1) às leis do país no qual se coleta a informação; 2) às leis do país do pesquisador, e 3) às leis ou acordos internacionais<sup>64</sup>. Além disso, os pesquisadores podem estar sujeitos às leis regionais internacionais, como as leis da União Europeia (UE), às quais está submetido o Programa DOBES nos Países Baixos; ou às leis que constam na Constituição para direitos autorais, caso dos programas DMA e AILLA nos Estados Unidos.

Apesar de cada país possuir um conjunto de leis que regulam os A) direitos de propriedade intelectual, B) os direitos de autor, C) os direitos morais e os D) direitos de acesso. As nações signatárias têm como referência também as resoluções da Convenção de Berna e as orientações de instituições como a *World Intellectual Property organization (WIPO)* e a *Trade Related Aspects of Intellectual Property Rights (TRIPS)* para temas correlatos (DWYER, 2006). Abaixo, veremos cada um desses direitos com maiores detalhes:

---

<sup>61</sup> Cf. Best Practices in Language Documentation. Disponível em: <http://emeld.org/school/index.html>. Acesso em: 04 set. 2021.

<sup>62</sup> Cf. Documentation of Endangered Languages. Disponível em: <<https://dobes.mpi.nl/dobesprogramme/>>. Acesso em: 04 set. 2021.

<sup>63</sup> Cf. AIATSIS. Code of Ethics. Disponível em: <<https://aiatsis.gov.au/research/ethical-research/code-ethics>>. Acesso em: 04 set. 2021.

<sup>64</sup> Cf. Consent, Rights and Intellectual Property: Navigating Language Documentation, Archiving, and Research. Disponível em: <<https://www.alaska.edu/colang2016/courses/workshop04/>>. Acesso em: 28 fev. 2023.

A) Em relação aos direitos de propriedade intelectual, a WIPO<sup>65</sup> (em português, “Organização Mundial de Propriedade Intelectual”) define propriedade intelectual como uma espécie de direito patrimonial que permite aos criadores ou titulares obterem os benefícios de seu trabalho, ou investimento em uma criação graças ao controle que lhes é dado sobre a utilização de sua propriedade. Essas “criações do pensamento” têm como representações livros, pinturas, interpretações musicais, filmes, poesias, folclore, etc. Esse conceito não possui equivalente na comunidade indígena, contudo há estudos sobre conhecimento e direito de propriedade indígena que podem ser aprofundados em algumas obras como HANSEN et al. (2003) e SULLIVAN (2002).

O TRIPS (em português, “Aspectos do Direito de Propriedade Intelectual Relacionados ao Comércio”), criado pela Organização Mundial do Comércio (OMC), regula o padrão de proteção mínima à propriedade intelectual e países signatários como o Brasil, desde 1994, são obrigados a revisar as leis nacionais com o fito de adaptá-las a esse padrão<sup>66</sup>. No país, a legislação que regula a propriedade intelectual é a Lei 9.279/96 e o órgão responsável pela concessão dos direitos de propriedade intelectual é o Instituto Nacional de Propriedade Industrial (INPI). Por outro lado, as regras que estipulam a proteção da obra são determinadas pela Lei 9.610/98. Conforme a Associação Brasileira de Propriedade Intelectual (ABPI)<sup>67</sup>: “Uma obra autoral não precisa ser registrada para estar protegida, mas o registro facilita a garantia dos direitos autorais para problemas futuros na esfera judicial”. A WIPO<sup>68</sup> ao tratar do tema propriedade intelectual garante a proteção jurídica da invenção ou criação da mente por um período de 50 anos a favor do inventor, criador ou titular do direito sobre a invenção.

A utilização de alguma forma dessas “criações do pensamento” ocorre por meio do consentimento informado que se refere à negociação entre o pesquisador e os colaboradores (falantes ou comunidade produtora da informação) sobre o futuro de todo o material, ou seja, quem terá acesso à informação, onde será armazenada, de

---

<sup>65</sup> Cf. World Intellectual Properties Organization. Disponível em: <<https://www.wipo.int/portal/en/index.html>>. Acesso em: 04 set. 2021.

<sup>66</sup> Cf. O que é propriedade intelectual? Disponível em: <<https://abpi.org.br/blog/o-que-e-propriedade-intelectual/>>. Acesso em: 04 set. 2021.

<sup>67</sup> Cf. Associação Brasileira de Propriedade Intelectual. Disponível em: <<https://abpi.org.br/blog/o-que-e-propriedade-intelectual/>>. Acesso em: 04 set. 2021.

<sup>68</sup> Cf. Word Intellectual Property Organization (WIPO). Disponível em: <[https://www.wipo.int/edocs/pubdocs/pt/sme/858/wipo\\_pub\\_858.pdf](https://www.wipo.int/edocs/pubdocs/pt/sme/858/wipo_pub_858.pdf)>. Acesso em: 03 set. 2021.

que forma e quem tomará futuras decisões a respeito de seu uso com implicações éticas e jurídicas. Essas discussões têm relação com preocupações relacionadas aos direitos de propriedade intelectual.

Acerca do tema, Epps et al. (2022, p. 10-11) afirma que: “[...] na maioria desses casos, os titulares dos direitos<sup>69</sup> originais retêm todos os direitos de propriedade intelectual e/ou cultural. Assim, podem conceder licenças não exclusivas ao arquivo e aos usuários do arquivo; os detalhes variam entre os arquivos e conforme as leis específicas de direitos autorais do país onde cada arquivo está localizado [...]”. Essa operação ocorre, por exemplo, no momento em que há um depósito de um recurso no AILLA, pois “[...] al no ser que se llegue a otro acuerdo por escrito, usted le da permiso a AILLA (le da licencia) para publicar el recurso en el internet, para hacer copias del recurso, para transferir el recurso a otro archivo y para utilizar el recurso para propósitos no lucrativos o académicos (como en una presentación sobre el archivo). El permiso no es exclusivo: usted también pasa por estos procesos y sólo usted puede derivar trabajos sin algún permiso explícito. La licencia o el permiso de AILLA es perpetuo”<sup>70</sup>.

A partir das informações acima, é possível concluir que o consentimento recebido pelo pesquisador e sua equipe para avançar com o projeto na pesquisa de campo já representa a divisão do direito autoral entre as partes envolvidas no processo de documentação. Seja o pesquisador representando os interesses do financiador do projeto ou o pesquisador defendendo os interesses particulares da própria pesquisa. A licença concedida ao programa de documentação, implicitamente inscrita no processo de uso do ALD, denota que este arquivo passa a ter privilégios de utilização dos recursos linguísticos digitalizados similares aos detentores originais dos direitos autorais. Entretanto, essa perda de exclusividade de direitos de autor não retira os direitos de propriedade intelectual dos criadores do pensamento (falante e/ou comunidade de fala) nos casos de cessão de consentimento ou licença. Em linha com essas reflexões, podemos inferir que pelo fato dos programas de documentação DOBES e DMA funcionarem abaixo do mesmo guarda-chuva jurídico das legislações

---

<sup>69</sup> Os mesmos autores afirmam que os titulares dos direitos são as pessoas com o direito de reivindicar direitos autorais sobre a “criação do pensamento” ou exigir direitos de propriedade intelectual, propriedade cultural, direitos morais, entre outros, cuja propriedade é garantida pela criação, associação, filiação ao grupo ou herança.

<sup>70</sup> Cf. Derechos de Propiedad Intelectual. Disponível em: < <https://ailla.utexas.org/es/site/derechos>>. Acesso em: 01 abr. 2023.

(nacionais e/ou transacionais) que versam sobre os princípios do tema Propriedade Intelectual, essas regulações também lhes sirvam.

Como afirmam Thieberger & Musgrave (2007), por tratar-se de questão complexa a ser explicada para uma comunidade onde a maioria das pessoas não possui a compreensão adequada do funcionamento do mundo externo, deve-se considerar possíveis dificuldades na atividade de campo (eletricidade, Internet e conflitos internos, etc.) como eventos que podem provocar a desistência dos colaboradores durante o desenvolvimento da pesquisa. Assim, deve ser plenamente admissível aos falantes, mesmo que tenha havido o consentimento antes, retirarem-se do projeto, a qualquer momento, caso desejem, com a exclusão de todo material coletado (direitos de propriedade intelectual). A importância do consentimento como boa prática no projeto de pesquisa nos remete às seguintes considerações.

Há três tipos de consentimento informado: O consentimento escrito, o consentimento verbal, e o consentimento de terceiros. Vejamos abaixo, mais informações sobre eles:

- O consentimento escrito tem como vantagem permitir a ambas as partes ter o registro do acordo, porém tem a desvantagem de exigir o anonimato dos colaboradores, algo que por vezes pode ser inadequado e gerar desconfiança, de modo que os pesquisadores geralmente têm alguma resistência a seu uso.
- O consentimento verbal, apesar da resistência da sociedade hegemônica, é o meio preferido de consentimento dos pesquisadores, considerando-se determinados contextos, como o das sociedades indígenas americanas, onde está provado que contratos verbais estreitam os vínculos e estabelecem maior relação de confiança de tal forma que os comitês institucionais de avaliação têm sido sensíveis em relação ao reconhecimento desse tipo de acordo.
- O consentimento de terceiros consiste no uso de um intermediário, por exemplo, um líder local para a negociar um contrato entre os participantes. A participação de um intermediário na confecção do contrato (escrito ou verbal) é considerada o modo mais apropriado para estabelecer um mínimo de confiança entre as partes e facilitar a comunicação do pesquisador com a comunidade de fala. Assim como acontece na legislação norte-americana, na maioria dos países não há leis de consentimento específicas, e sim genéricas, as quais não protegem os produtores da informação (AUSTIN, 2010; DWYER, 2006).

B) Os direitos de autor referem-se à propriedade e à distribuição de uma obra específica. Como é uma forma de propriedade, podem ser herdados, presenteados ou vendidos. Na lei dos direitos autorais, em perspectiva monetária, se o direito do autor for violado, este perde o valor correspondente. Apesar do enfoque pecuniário não importar aos projetos de documentação linguística, pois não se trata de atividade lucrativa, eles estão sujeitos a essas leis. Os direitos autorais variam conforme os tipos de materiais e os países, por exemplo, no Reino Unido, há quatro categorias, com diferentes dispositivos legais: obra literárias (livros, impressos), gravações de som, banco de imagens, filmes. Conforme Newman (2007, p. 30), os direitos autorais têm como características fundamentais: dar ao autor o monopólio sobre a obra; limitar a terceiros a utilização não autorizada da obra; garantir automaticamente a proteção dos direitos do autor; ser transferível por venda, doação, herança; possuir duração longa, etc. Na legislação internacional, há ainda os direitos autorais exclusivos, em que somente o editor tem o direito de copiar e distribuir o livro, e os não exclusivos, situação em que os direitos autorais são compartilhados entre o autor e outras entidades (DWYER, 2006; NEWMAN, 2007).

C) Os direitos morais representam os direitos não econômicos sobre determinada obra. Conforme a Convenção de Berna<sup>71</sup>, estão garantidos ao produtor da informação, a quem narra a história (falante/consultores), alguns direitos legais sobre sua obra. Como a força jurídica dos direitos morais não foram testadas empiricamente, os produtores de dados (falantes/consultores) e os pesquisadores preservam alguma proteção através de seus direitos econômicos. Sendo assim, Conforme Austin (2010, p. 44), mesmo com a venda dos direitos de publicação, o uso indevido do material pelo editor acarretará a violação dos direitos morais do autor.

D) Os direitos de acesso e uso se referem à licença dada pelos falantes para o pesquisador depositar o corpus de dados primários coletado em arquivo linguístico vinculado ou não a um programa de documentação (por exemplo, DOBES, DMA ou AILLA) pelos anos ou décadas seguintes para o acesso de terceiros (comunidades de fala, estudantes, pesquisadores e público em geral). Conforme a experiência dos pesquisadores, o período em que ainda estão no campo é o momento adequado para

---

<sup>71</sup> Cf. Direitos econômicos: “Independentemente dos direitos econômicos do autor e mesmo após a transferência de tais direitos, o autor terá o direito de reivindicar a autoria da obra e se opor a qualquer distorção, mutilação ou outra modificação, ou outra ação derogatória em relação à referida obra, o que seria prejudicial à sua honra ou reputação.” Disponível em: < [http://www.wipo.int/treaties/en/ip/berne/trtdocs\\_wo001.html](http://www.wipo.int/treaties/en/ip/berne/trtdocs_wo001.html) >. Acesso em: 03 set. 2021.

se conseguir a permissão de gravação com os falantes e demais participantes. Nesse caso, algumas questões geralmente levam a debates éticos, como a relação entre a coleção e o acesso público. Quanto ao ALD, existem formas de coordenar os direitos relacionados ao material depositado no arquivo e os direitos de obter acesso e/ou usar o referido material. Vários arquivos utilizam um sistema conhecido como “protocolo” (AUSTIN, 2006; DWYER, 2006) que permite o acesso controlado, como o utilizado pelo AILLA que consiste em três níveis, como: 1 - “completamente aberto” (qualquer pessoa tem acesso aos materiais); 2- “completamente fechado” (somente o depositante tem acesso aos materiais); 3- “parcialmente aberto”; nesse caso, o acesso ao material está sujeito a condições específicas determinadas pela (o): (i) restrição imposta pelo falante gravado; (ii) tipo de material (gênero, religiosidade); tipo de usuário (gênero, associação tribal ou etnia).

O *Endangered language Archives* (ELAR), utiliza os formulários que acompanham todos os depósitos de materiais e consegue especificar claramente os direitos de acesso (NATHAN, 2010 apud AUSTIN, 2015). Até mesmo os repositórios digitais particulares devem encontrar meios de tornarem-se acessíveis à comunidade nativa, principalmente onde não há Internet, mas há computadores e outras mídias eletrônicas.

Para finalizar, ressaltamos que as atividades dos programas de documentação (DOBES, DMA e AILLA) seguem regulamentações locais, regionais e/ou internacionais. Sempre que a legislação nacional for insuficiente para resolver problemas de direitos (propriedade intelectual, autorais, morais, etc.) os acordos e convenções internacionais deverão ser acionados para temas controversos. Como vimos, há organizações para tratar temas específicos como a WIPO e TRIPS voltados à discussão sobre propriedade intelectual. Reforçamos a importância que os diversos tipos de consentimento têm para as boas práticas de documentação e, posteriormente, para a publicação do resultado do projeto; bem como realçamos o papel de cada um dos direitos no processo de documentação, com destaque para a complexa questão de quem é o verdadeiro detentor da propriedade intelectual daquilo que é produzido na atividade de documentação. Para responder a essa dúvida lançamos, além do estudo de Epps et al. (2022), a análise do Programa AILLA no processo de arquivamento de recurso no ALD em que este, por meio da aceitação tácita do conteúdo do contrato pelo usuário, divide os direitos autorais com o depositante e/ou comunidade de fala, o que não configura nenhuma ilegalidade. Os

direitos adquiridos passam a permitir que o arquivo tenha uma licença (perpétua) com os mesmos privilégios dos donos originais dos direitos autorais, o que lhe permite utilizar (conforme o que delimitam as leis locais e internacionais para direitos autorais) o conteúdo dos recursos publicados. Devido às limitações do escopo desta pesquisa, fizemos uma análise superficial sobre este tema nos programas DOBES e DMA, concluindo que procedimentos semelhantes são adotados com as mesmas finalidades.

#### 2.4.2.4 – Pesquisa Colaborativa

Em termos de pesquisa colaborativa, há dois fatores cruciais para o êxito no trabalho de campo linguístico: uma boa relação entre os pesquisadores e colaboradores locais e a elaboração de um plano de trabalho bem organizado, baseado no conhecimento compartilhado e na negociação conjunta dos objetivos (ver anexo, Figura I). Nesse caso, resta aos pesquisadores que não possuem relação com uma comunidade de fala, construí-la cooperativamente. Para os que se encontram na situação oposta, a cada novo projeto é preciso retomar as conexões, protocolos e renegociar os papéis dos participantes. Com esse objetivo, em suma, listamos as etapas a percorrer que envolvem: 1) os critérios relacionados ao local associado à variedade da língua como diversidade linguística (tipologia ou variedade), conveniência política (controle sobre o deslocamento do pesquisador na comunidade), conveniência logística (limitar ou não o espaço geográfico da pesquisa), conveniência interpessoal (evitar o conflito de interesses entre os pesquisadores); 2) o contato de consultores nativos (de todos os contatos, o mais importante), acadêmicos (da região são ajuda importante para impulsionar a pesquisa), funcionários (importantes para obtenção de permissão de pesquisa), figuras locais (falantes não nativos fornecem perspectivas étnicas sobre o povo pesquisado); 3) o estabelecimento de relações de confiança antes de acessar a comunidade, o trabalho cooperativo entre consultores e pesquisadores, a constituição de equipes de pesquisadores da comunidade; 4) a compensação que deve ser feita pelos pesquisadores aos consultores e à comunidade pela contribuição à pesquisa pode ser materializada na forma de gravações de áudio e vídeo, ortografias práticas, livros, valores monetários (nesse caso, a sugestão é dar o mesmo valor a todos os participantes). Veja mais sobre pesquisa colaborativa em Bower (2015), Schultze-Berndt (2015) e Dwyer (2006).

Por fim, em resposta às questões colocadas no início desta seção, acreditamos que a utilização de um padrão normativo único em pesquisa de campo, sempre desafiadora, ajuda porque orienta o pesquisador frente a seus objetivos de trabalho básico, mas isso não significa que ele deve ignorar seus conhecimentos empíricos para superar situações não previstas que por ventura possam ocorrer. Enfim, os padrões normativos éticos norteiam os rumos das atividades com os colaboradores e na relação com a comunidade de fala, mas são insuficientes para as complexas situações que surgem durante o andamento do trabalho.

### 2.4.3 – Aspectos linguísticos

#### 2.4.3.1 – Tipologia

A tipologia linguística é a subárea da linguística responsável pelo estudo e classificação das línguas conforme suas características estruturais. Sua finalidade é a descrição e explicação da diversidade estrutural e propriedades comuns das línguas existentes no mundo. Nesses parâmetros, entre outras, se enquadram as línguas indígenas brasileiras, principalmente em critérios como organização de sistemas de sons e estrutura gramatical, constituindo, conforme Seki (2000, p. 239), tipos linguísticos únicos presentes em algumas línguas como: Tikuna, Irantxé/Münkü, Trumái, Máku, Aikaná, Arikapú, Jabuti, Kanoê, Koaiá, etc.

Em linha com essa realidade, um projeto de documentação linguística deve considerar o trabalho com as línguas de maior diversidade tipológica possível, algo que deve exigir, como afirma Wittenburg (2002), uma estrutura de arquivamento flexível que reflita no design da arquitetura de dados, como os bancos de dados lexicais que facilitam a pesquisa de morfemas nas línguas polissintéticas, além de atender futuras demandas próprias da pesquisa dos falantes nativos. Como os projetos devem colaborar estreitamente com as comunidades de fala, os arquivos também devem corresponder às necessidades desses grupos sociais.

#### 2.4.3.2 – Tipos de dados

Conforme Wittenburg et al. (2002), associada à preocupação com aspectos tipológicos, a documentação linguística deve conter dados dos seguintes tipos: (1) curta descrição da filiação genética da língua acompanhada das principais características tipológicas; (2) resumo do contexto sociolinguístico (número de falantes, nível de multilinguismo na comunidade, circunstâncias da documentação, histórico de pesquisa, etc.); (3) corpus de gravações de áudio e vídeo (se possível) com diferentes manifestações da língua (mitos, anedotas, conversas casuais, debates políticos, eventos de fala, etc.) com uma transcrição, uma tradução, anotações e comentários linguísticos; (4) descrição dos metadados, por sessão gravada, em formato padronizado com informações sobre data, local, responsável pela coleta, equipamento utilizado, identificação dos falantes, etc.; (5) descrição fonética ou fonológica detalhada que contemple um inventário das distinções fonêmicas segmentadas e prosódicas; e as regras alofônicas / alotônicas; e a (6) descrição das convenções da transcrição e, se for o caso, a descrição das convenções ortográficas; a lista explicativa com todas as abreviações e os símbolos presentes na documentação. Esse corpus de gravações associado às suas transcrições e traduções representa o núcleo da documentação linguística que pode ser acompanhado de um esboço gramatical, um dicionário bilíngue e ensaios linguísticos ou etnográficos descritivos (SCHULTZE-BERNDT 2015; BOWERN 2015).

#### 2.4.4 – Prática documental

Como vimos, a função dos dados na documentação linguística não tem equivalente na linguística descritiva. Enquanto nesta, o mais importante é a produção de gramáticas e dicionários, onde os dados primários servem basicamente para apoiar a análise e descrição do linguista, como produtos secundários que derivam desses dados (AUSTIN, 2006); na documentação linguística, o principal objetivo é o registro (coleta), a transferência, o processamento, o arquivamento e a mobilização dos dados, permitindo uma utilização não restrita ao meio acadêmico, mas também acessível aos membros da comunidade onde ocorreu a documentação. Conforme Woodbury (2003, p. 46-47), um corpus documental bem formatado deve ser: 1) Diversificado: contendo amostras de uso da língua em vários gêneros e contextos socioculturais, incluindo

dados elicitados; 2) Amplo: graças aos recursos de armazenamento e manipulação da moderna Tecnologia da Informação e Comunicação (TIC), um corpus digital pode ser extenso e incorporar mídia e texto; 3) Contínuo, distribuído e flexível: os dados podem ser adicionados ao corpus a partir de quaisquer fontes disponíveis e expandidos quando novos materiais estiverem disponíveis; 4) Transparente: o corpus deve ser estruturado de modo a ser utilizável por pessoas que não os compilaram, incluindo futuros pesquisadores; 5) Conservável e portátil: preparado para permitir que seja arquivado para preservação a longo prazo e não se restrinja a ambientes específicos de TIC; 6) Ético: gravado, transferido, processado, arquivado e mobilizado com a devida atenção aos princípios éticos (vistos anteriormente) de acordo com todos os protocolos necessários para o acesso e uso.

Conforme Austin (2010, p. 19), o processo de documentação linguística depende do desenvolvimento de um projeto de pesquisa voltado ao trabalho com comunidades de fala. Tal atividade inserida no bojo das atividades de campo envolve a coleta de dados, ou seja, o estudo da forma como a língua é falada em seu ambiente natural. Nesse contexto, como vimos, a partir de considerações éticas, teóricas e metodológicas, o pesquisador interage com os falantes de modo que o fluxo das suas tarefas lhe permita desenvolver as suas atividades. Conforme Austin (2015), alguns projetos, nas primeiras décadas do século XX, como o DOBES (2001), o E-MELD (2001-2006), o *Committee on Endangered Languages and Their Preservation* CELP (2007)<sup>72</sup> e outros criaram modelos próprios de documentação linguística a serem adotados pelas pesquisas que financiam. Tais iniciativas especificam detalhes que servem para o armazenamento de materiais; formatos de gravação e análise; ferramentas de software para anotação de áudio e vídeo (p. ex., ELAN) e criação e gerenciamento de metadados (p. ex., IMDI). Esses princípios gerais ou “melhores práticas”<sup>73</sup> também se estenderam à prática documental (gravação, transferência, processamento, arquivamento e mobilização da sua corpora de dados) e à tipologia dos tipos de dados (HIMMELMANN, 2012; LUPKE, 2009; SEIFART, 2008 apud AUSTIN, 2015). A seguir, apresentamos um breve relato técnico desses procedimentos a partir das práticas dos três protagonistas nesse processo:

---

<sup>72</sup> Cf. Committee on Endangered Languages and Their Preservation (CELP). Disponível em: <<https://www.linguisticsociety.org/about/who-we-are/committees/endangered-languages-and-their-preservation-celp>>. Acesso em: 19 set 2021.

<sup>73</sup> Para maior aprofundamento sobre as discussões acerca das “melhores práticas” na documentação linguística, conferir Austin (2015); Lupke (2009), Seifart (2008), Nathan (2004).

depositantes, usuários e arquivistas.

#### 2.4.4.1 – Processamento linguístico

Cada etapa do processamento linguístico trata, conforme Austin (2007), do momento em que se “agrega valor” aos registros por meio da análise linguística (transcrição, tradução e anotação) e vinculação dos metadados às gravações. Assim, o trabalho de processar o material oriundo da gravação e captura é uma das tarefas que exige maior tempo do pesquisador. Se por um lado, a transcrição e a anotação do áudio podem levar horas (conforme o domínio da língua e o detalhamento da anotação), o mesmo procedimento é ainda mais demorado e árduo no caso de vídeos (nesse caso, são necessários recortes que permitem a criação de blocos de mídia de tamanho gerenciável).

A análise linguística desses registros deve ser apoiada na estruturação teórica definida pelo pesquisador e por sua representação nos metadados que acompanham os materiais selecionados. Em vista disso, Austin (2006, p. 95) afirma que:

Existem boas razões para visar um certo grau de padronização ao processar os materiais, incluindo transparência, portabilidade e facilidade de compartilhamento e acesso (Bird & Simons 2003). A transcrição fonética deve seguir as convenções da Associação Fonética Internacional (IPA), e a transcrição fonêmica deve ser IPA ou um padrão regionalmente reconhecido. Tags de anotação gramatical (ou seja, os rótulos abreviados para, por exemplo, parte das categorias de fala) deve seguir a prática linguística geral, por exemplo, as recomendações do EUROTYP ou E-MELD (incluindo sua ontologia), com uma lista de abreviações e símbolos relevantes fornecidos como metadados [...]

Em termos de processamento para os diferentes tipos de dados, Austin (2007, p. 121) aponta a existência de vários tipos de codificação e armazenamento, entre os quais destacam-se:

- Codificação de caracteres (formato de character): windows, ansi, unicode, utf-8, big5, jisc, etc.

- Codificação de dados (padrões de estruturas de marcação): padrão XML, Shoebox, Toolbox, Elan, Tabela Word, etc.
- Codificação de material (empacotamento de dados): rtf, doc, docx, pdf, xls, etc.
- Armazenamento físico (mídias de armazenamento): hd, dvd-rom, cd-rom, minidisco, dat, pendrive, flash memory, etc..

Em pesquisa rápida na Internet é possível verificar a existência de grande variedade de softwares voltados ao processamento linguístico, porém a escolha destes deve possibilitar, conforme Bower (2015, p. 53), o acesso a três recursos fundamentais: (1) a localização rápida e precisa dos dados no computador; (2) a busca refinada de aspectos específicos nos dados; (3) a formação de banco de dados capaz de produzir relatórios para a comunidade de fala e armazenar textos convertidos e de gramática de referência. Na mesma obra, a autora traz um manual completo sobre as melhores práticas no processo de análise linguística, as quais não detalharemos devido às limitações de escopo desta pesquisa.

Finalmente, o linguista após cumprir as etapas de gravação e transferência deve dar andamento a sua pesquisa com o processamento linguístico. Para isso, deverá proceder com as atividades de análise. Em vista disso, abaixo, são listadas indicações de alguns softwares<sup>74</sup> básicos, acompanhados de suas respectivas finalidades<sup>75</sup>:

- Filemaker, Access, Base: Organizar anotações de campo e gravações.
- Audacity, Praat, Quicktime, VLC: Processar gravações (áudio e vídeo).
- ELAN, TranscriberAG, Transana, Clan: Transcrição.
- Toolbox, ELAN, Tlex, TextStat: Criação, gerenciamento e pesquisa de corpus.
- Toolbox, Flex: Análise / interlinearização.
- Praat, Toolbox, Flex, TextStat: Análise / marcação.
- Lexique Pro, Toolbox, TshwaneLex, WeSay, Lexus: Confecção de léxico / dicionário.
- Word, LibreOffice: Escrita de materiais secundários (artigos de jornal, gramática de referência, dissertação, materiais da comunidade).
- Cuped: Publicação na web.

<sup>74</sup> Cf. Software tools for language documentation (1). Disponível em: <<https://slideplayer.com/slide/3199792/>>. Acesso em: 19 out. 2021.

<sup>75</sup> Cf. Software tools for language documentation (2). Disponível em: <<https://slideplayer.com/slide/3172618/>>. Acesso em: 19 out. 2021.

- Nuvem (Dropbox, Google Drive, Sugarsync, OneDrive, iCloud, etc.); Serviços de backup da universidade; Mídia física (HD externo, DVD, Pendrive, etc.): Ferramentas de backup<sup>76</sup>.

A lista de softwares disposta representa uma sugestão que deve garantir o fluxo de trabalho (ver exemplo de fluxo de trabalho para documentação linguística no Anexo B) do pesquisador, mas é importante ressaltar que uso desses softwares juntamente com os equipamentos escolhidos dependerá das escolhas do pesquisador. Alguns dos softwares apontados estão no formato proprietário<sup>77</sup>, ou seja, armazenam dados no formato que não é ideal para armazenamento e acesso a longo prazo.

Conforme Trilsbeeck & Wittenburg (2006), no campo geralmente o pesquisador precisa lidar com diferentes dificuldades que exigem flexibilidade e rapidez em decisões que determinarão a escolha dos instrumentos necessários (hardware e software) que não onerem o tempo de trabalho e a produtividade da pesquisa.

Como vimos, o resultado do trabalho do pesquisador-depositante tem relação direta com os recursos e ferramentas empregados, associados ao grau de eficiência de sua aplicação na tarefa de documentação linguística, de modo que seus resultados sejam satisfatórios quando apresentados à comunidade indígena e à comunidade científica no ato de sua publicação.

#### 2.4.4.2 – Arquivo digital

Os arquivos tradicionais, voltados ao armazenamento de objetos físicos, possuem algumas limitações se comparados aos arquivos digitais. Os primeiros têm a limitação de acesso aos objetos expostos por questões de espaço físico, ameaça de depredação, impossibilidade de criar cópias, etc. Assim, as melhores soluções envolvem a construção de locais apropriados e custosos para a preservação dos acervos. O arquivo digital moderno, por outro lado, rompe essas barreiras, pois sua prioridade é a sobrevivência do corpus de dados primários em servidores (computadores dedicados ao armazenamento de dados), de modo que preservem seus requisitos básicos, como edição e acessibilidade para os usuários (depositantes, internautas ou

---

<sup>76</sup> Conforme Sawaya (1999, p. 41), o backup envolve a criação de “um arquivo auxiliar, imagem do arquivo fonte, utilizado como base de recuperação de dados quando da ocorrência de um defeito ou perda de dados”.

<sup>77</sup> Cf. O software livre é gratuito? Disponível em: <<https://canaltech.com.br/software/O-software-livre-e-gratuito/>>. Acesso em: 21 out. 2021.

arquivistas) sem a necessidade de despesas com um espaço físico fixo e centralizado.

Existem dois diferentes tipos de materiais digitais que armazenam documentos linguísticos: (1) os materiais digitais de dados que se referem às gravações anotadas, léxicos, esboços de gramática, notas de campo, descrições fonéticas, metadados, ontologias e informações geográficas; e (2) os materiais digitais de mídia tradicionais como texto (doc, docx, txt, odt, etc.), áudio (mp3, aac, wma, flac, etc.), vídeo (avi, mov, mkv, avchd, etc.) e imagem (tiff, png, jpeg, raw, etc.). Conforme Trilsbeek & Wittenburg (2007), os acervos devem permitir não somente o armazenamento de informações, mas também o acesso e a modificação das mesmas. Algo que, ocasionalmente, pode causar problemas do ponto de vista do arquivamento, tendo em vista que esses tipos de arquivos têm duas funções: o armazenamento a longo prazo e o acesso a curto prazo; escopos que atendem aos interesses de diferentes grupos envolvidos na documentação linguística, entre os quais destacam-se o depositante, o usuário e o arquivista, analisados a seguir.

#### 2.4.4.3 – Depositante

A responsabilidade do depositante no processo de desenvolvimento da pesquisa documental envolve diversos participantes (consultores, estagiários, comunidade de fala, instituições de ensino, etc.), e engloba várias etapas que compõem sua metodologia, entre as quais, destacamos as seguintes: 1) gravação, 2) transferência e 3) processamento. Esses procedimentos são executados, como vimos, durante as atividades de campo, com ferramentas digitais e computacionais para captar as principais manifestações culturais e peculiaridades linguísticas de uma comunidade de fala em meio às suas múltiplas práticas existenciais cotidianas.

1) Sobre a atividade de gravação, durante a pesquisa de campo, Margetts & Margetts (2014, p. 14) afirmam que seu objetivo principal para a documentação linguística “é registrar a fala para analisar a estrutura da língua”. Poderíamos complementar essa constatação, afirmando também que o processo de gravação é etapa essencial no registro duradouro e multifuncional de aspectos linguísticos e culturais de uma comunidade de fala, em que a língua pode estar ou não ameaçada de extinção.

O aspecto empírico dessas reflexões se reflete na compilação das recomendações de Bowern (2015) e Margetts & Margetts (2014) para os

equipamentos indispensáveis no registro do corpus de dados primários durante a pesquisa de campo (ver Anexo A). Por representar um elemento estruturante da documentação, a gravação de áudio e vídeo deve seguir os parâmetros propostos pelo método de registro documental – que abrange procedimentos básicos e lista de softwares -, amplamente empregado por linguistas e documentadores (ver anexos C, D e E), como Trilsbeek & Wood (2008)<sup>78</sup>.

Entre os procedimentos básicos assinalados, são consideradas as observações sobre melhores práticas durante as gravações e registro do corpus documental. Nesse caso, é possível encontrar variadas proposições de instituições envolvidas com documentação, algumas das quais citadas nesta dissertação. Entretanto, o padrão DOBES destaca-se por ser completo, testado e orientado por guias e tutoriais. Para complementar essas orientações, Austin (2007) sugere que cada parte, dessa etapa da pesquisa, seja dividida entre os membros da equipe de documentação, de modo que as gravações de áudio sejam efetuadas separadamente das gravações de vídeo, e, obrigatoriamente, acompanhadas de anotações de campo, da rotulagem das mídias e de metadados. A reunião desses materiais forma os dados primários da pesquisa.

2) Em relação à transferência, Austin (2010) afirma que se trata de manipulação ou deslocamento dos materiais coletados e acumulados durante o período da pesquisa de campo, de locais de armazenamento temporário e inseguro (disco externo, pendrive, memórias flash, cd, dvd, fita cassete, etc.) para artefatos e locais de armazenamento de dados definitivos e confiáveis, como PCs ou Notebooks robustos, com processadores e HD de alto desempenho e capacidade. Nessa etapa, deve-se fazer, por exemplo, a conversão das gravações analógicas ou dos textos escritos em papel para o formato digital, entre outros procedimentos.

3) Na fase de processamento, portanto, fora do ambiente de pesquisa de campo, o pesquisador-depositante deve finalizar o processo de conversão das notas de campo, escritas em papel, para o formato digital, e transferi-las para o computador, o mais rápido possível, para que não se percam as notas, abreviações e comentários. Os dados primários e suas pastas de arquivamento devem ser nomeadas e identificadas de modo coerente e padronizado conforme seu conteúdo (para mais informações, ver MOORE, 2011). Os mesmos procedimentos para o armazenamento

---

<sup>78</sup> Cf. MPI for Psycholinguistics (TRILSBEEK & WOOD, 2008). Disponível em: <<https://www.mpi.nl/corpus/a4guides/a4-guide-TMPGEnc.pdf>>. Acesso em: 25 ago. 2021.

e nomenclatura devem ser reproduzidos para os metadados. Nesse caso, deve-se ressaltar a importância da utilização de nomenclatura que facilite sua leitura conforme o parâmetro estruturado das tabelas relacionais de banco de dados (JOHNSON, 2004). Após a conclusão de todas as etapas do processamento, mesmo que a pesquisa continue em andamento, o pesquisador-depositante pode inserir os (primeiros) dados primários produzidos, com os metadados, no respectivo arquivo linguístico digital.

#### 2.4.4.4 – Usuário

O crescimento da quantidade de instituições envolvidas com pesquisa em documentação linguística tem levado ao aumento do número de pessoas que buscam informações sobre coleções documentais. O fato ocorre nos espaços acadêmicos, nas instituições de pesquisa ou em ambientes virtuais. Os grupos sociais que lideram esse interesse por conhecimento nos arquivos linguísticos são formados, geralmente, por estudantes, professores, jornalistas, comunidades de fala e público em geral.

Trilsbeek e Wittenburg (2006) identificam o interesse variado de cada um desses grupos nas pesquisas às bases de dados, de modo que defendem a necessidade da criação de formas diferenciadas de apresentação e disponibilização dos registros. Assim, propõem, por exemplo, a necessidade de distinguir a especificidade da consulta de três perfis de usuários diferentes: (1) um pesquisador interessado em localizar fenômenos estruturais linguísticos singulares de uma língua; (2) a pesquisa feita por um membro de comunidade indígena voltada à investigação da gravação de um ritual ou discurso descritivo das características culturais de seu povo; (3) e os pesquisadores de diferentes áreas do conhecimento que utilizam critérios de pesquisa diferentes da Linguística.

Como não é possível prever como os dados primários serão utilizados no futuro, Himmelmann (2006) afirma ser indispensável que eles sejam fartamente acompanhados de metadados, de modo que sejam plenamente acessíveis à ampla variedade de usuários e interesses. Além disso, as dificuldades futuras no acesso ao banco de dados linguístico podem ser minimizadas por intermédio da padronização do modo pelo qual o material é coletado e armazenado. Nesse sentido, é importante que: (1) o material seja da mais alta qualidade possível diante do estágio atual da tecnologia; (2) o material deve ser claramente organizado e documentado; (3) todos

os objetos devem ser acessíveis, como recursos individuais, da forma mais neutra de representação, permitindo aos usuários combiná-los do modo mais conveniente; (4) todas as decisões sobre padrões de codificação, formatos de material, etc., têm que ser documentadas (TRILSBEEK & WITTENBURG, 2009; AUSTIN, 2007).

#### 2.4.4.5 – Arquivista

Como sabemos, uma documentação linguística, adequadamente elaborada, precisa ser ampla, contínua, sem custo, verificável, ética, preservável a longo prazo e portátil em linhas gerais. De modo que forneça farto material de áudio / vídeo, transcrição e anotação associados a metadados contextuais. Tal registro tem várias finalidades, entre as quais enfrentar o desafio das línguas ameaçadas, fornecer dados estruturais linguísticos, crescer e melhorar a base científica e economizar os recursos de pesquisa (WOODBURY, 2006; AUSTIN, 2013). Em torno disso, o arquivamento tem papel crucial na sustentação e preservação dos dados primários.

Os pilares que sustentam a qualidade do material coletado e de sua análise estão diretamente relacionados à qualidade de sua organização e documentação. Esses atributos devem estar associados a um sistema de catálogo e estrutura das informações, na forma de metadados, geralmente representado pelo padrão IMDI, responsável por operações como depósito, gerenciamento e acesso ao arquivo de dados primários.

Entre as principais preocupações do arquivista está a manutenção dos registros documentais no longo prazo, de modo que as informações relacionadas às práticas e estruturas linguísticas fiquem disponíveis e acessíveis indefinidamente (TRILSBEEK & WITTENBURG, 2006). Essa perspectiva de trabalho tem como contraponto e dilema o trato com as demandas de curto prazo no que tange à relação com o acesso aos dados primários envolvendo depositantes e usuários em geral.

A rotina de trabalho do arquivista responsável pela documentação linguística não se resume a sua perspectiva de preservação dos dados a longo prazo. Sua dimensão de tempo para o tratamento dos dados primários por vezes entra em conflito com as demandas de depositantes e usuários. As principais críticas feitas pelos arquivistas em relação aos documentos entregues para arquivamento estão na (s): 1) gravações em formato compacto de mídia com baixa qualidade de áudio e/ou vídeo; 2) utilização de softwares comerciais proprietários; 3) representação do material no

formato HTML<sup>79</sup>.

Em relação a cada um desses pontos, respectivamente, os arquivistas apontam a priorização das gravações com alta qualidade, não compactadas, pois esse formato permite a manutenção da compatibilidade das gravações para uso no futuro distante, em outro ambiente tecnológico. Acerca dos softwares cujo código-fonte não é aberto, a principal crítica se refere às alterações não controladas pelo usuário e a não limitação a entrada de conteúdo e a estrutura de dados. Por fim, a apresentação dos dados no formato HTML não garante expressamente a apresentação da estrutura e da organização documental<sup>80</sup>.

De acordo com Trilsbeek & Wittenburg (2007), o trabalho com os diferentes formatos das mídias documentais e a forma de sua apresentação ao arquivista tem relação direta com o tempo que depositantes e usuários despendem em termos de acesso às mídias que compõem a documentação linguística. Se por um lado, a apresentação e a consulta aos dados envolvem metas de curta prazo, por outro, as tarefas de armazenamento voltadas a preservação desses registros no longo prazo envolvem os requisitos exigidos pelos arquivistas já apresentados aqui. Ou seja, os conflitos entre as partes ocorrem pelas divergências entre os que priorizam a rápida apresentação e consulta aos dados (curto prazo) e aqueles que exigem formatos que atendam às exigências de conservação das mídias armazenadas, seguramente, para a utilização das gerações futuras (longo prazo).

A tarefa de garantir o êxito no armazenamento dessas mídias, no longo prazo, não exige do arquivista, unicamente, a tarefa de cobrar de depositantes formatos de documentos digitais que sejam compatíveis com essa pretensão. Há outras duas tarefas que precisam caminhar juntas àquelas já mencionadas: 1) garantir por tempo indeterminado o fluxo de bits (padrões binários armazenados na mídia) sem que estes sejam ameaçados pela vulnerabilidade dos dispositivos que os armazenam (hd, dvd-rom, cd-rom, memórias-flash, etc.) e 2) garantir a interpretabilidade e preservação das informações contidas nesses fluxos de bits, preservando sua estrutura. Infere-se, então, que o segundo objetivo somente será possível se forem desenvolvidos métodos que decodifiquem os dados contidos na forma de fluxo de bits (TRILSBEEK & WITTENBURG, 2007).

No que se refere à vulnerabilidade dos dispositivos de armazenamento, é

---

<sup>79</sup> Ibid.

<sup>80</sup> Ibid.

preciso distingui-los quanto ao tempo de duração. Os discos rígidos duram em média de quatro a seis anos. Mídias, como cd-rom e dvd-rom, têm duração um pouco maior, porém isso vai depender do modo como são conservadas; outras mídias semelhantes de armazenamento têm duração semelhante. Além desses, outro fator que reduz o ciclo útil de utilização dessas mídias é o processo de inovação tecnológica que limitará, daqui a algumas décadas, a somente algumas instituições especializadas a capacidade de ler o conteúdo das mídias obsoletas e processar as informações contidas. Assim, o que há de comum nessas mídias é que todas são incompatíveis com a necessidade de armazenamento a longo prazo desejada pelos arquivistas<sup>81</sup>.

Em torno das dificuldades de armazenamento a longo prazo assinaladas, algumas soluções estão sendo estrategicamente empregadas, entre as quais destacam-se duas: a primeira envolve a transferência contínua dos dados para as mídias tecnologicamente mais recentes, próprias para guardar volumes de dados que ultrapassam as unidades máximas de kilobytes utilizadas atualmente<sup>82</sup>; a outra solução, largamente utilizada, expõe a necessidade da distribuição dos dados armazenados em diversas cópias (backup). Ressalte-se que tais procedimentos devem ser implementados antes que a tecnologia de armazenamento utilizada se torne completamente insegura e obsoleta. Repositórios de documentação linguística de grande porte costumam utilizar essas estratégias para a preservação de seus dados.

Juntamente com a proteção dos fluxos de bits, algumas medidas devem ser tomadas para que a interpretabilidade e a legibilidade das informações sejam viáveis no longo prazo. Em vista disso, Trilsbeek & Wittenburg (2006, p. 320-324) apontam quatro procedimentos a serem executados acerca de: 1) como codificar corretamente as mídias (caracteres, imagens, sons e vídeos); 2) como estruturar texto e formato de material; 3) como estruturar e organizar um corpus; 4) como compatibilizar a capacidade de armazenamento do arquivo com o material enviado. Abaixo, traçamos um breve resumo de cada uma dessas etapas.

---

<sup>81</sup> Ibid.

<sup>82</sup> Conforme a StorageOne: “Cada byte corresponde a apenas um caractere entre as 256 combinações possíveis de 0 e 1. Nessa mesma linha, apesar do prefixo “kilo” representar mil unidades, um kilobyte é o agrupamento de 1024 bytes, dado que a computação é baseada no código binário; por isso, cresce exponencialmente. Na prática, 2 elevado à décima potência ( $2^{10}$ ) é igual a 1024. Sendo assim, 1.024 kilobytes representam um megabyte e assim em diante até chegar à unidade máxima yottabyte, equivalente a 1.024 zettabytes.” Disponível em: <<https://www.sto1.com.br/blog/armazenamento/unidades-de-armazenamento-de-dados>>. Acesso em: 29 out. 2021.

1) Para a codificação adequada das mídias, como as de áudio e vídeo, por exemplo, é importante que as diferentes informações contidas sejam armazenadas no padrão de bits (cuja perspectiva de uso é considerada indefinida) e empacotadas (codificadas) em formatos (mpeg2, mpeg4, mp3, avi, wma, etc.). Para que esses materiais possam ser desempacotados (decodificados), corretamente, daqui a dezenas ou mesmo centenas de anos (longo prazo) é fundamental que seus métodos de codificação sejam documentados, mediante metadados, conforme as orientações expostas a seguir:

- A documentação deve tratar do software (codificação / decodificação) e do tipo de codificação utilizada de modo simples.
- A codificação das mídias deve ser, prioritariamente, no padrão direto digital linear (alta definição) para os formatos de áudio (pcm, aiff, wav, etc.), imagem (tiff, png, raw, etc.), e vídeo (mpeg2, mpeg4, avchd, etc.).
- No arquivamento, deve-se evitar mídias de formato compactado (baixa definição).
- No geral, a proposta de preservação da documentação a longo prazo engloba os pontos assinalados considerando que os softwares padrão responsáveis pela codificação não devem ser proprietários.

2) Os procedimentos para a estruturação do texto e do formato dos materiais visam a identificação e a documentação completa dos componentes estruturais das unidades de interpretação (palavras, glossários, anotações, discursos, etc.). Em vista disso é importante:

- A utilização de tags (para identificar componentes do texto), e do xml (para descrever a estrutura do texto).
- A utilização de esquemas xml é fundamental para o controle das estruturas dos documentos e o gerenciamento das correções sofridas pelos materiais.
- A estrutura dos documentos tem relação direta com o formato dos materiais e este define como as informações são empacotadas (fluxo de bits).
- A garantia da interpretabilidade das informações empacotados no futuro (distante) é fundamental para a documentação do formato dos materiais.

3) A sistematização de um corpus documental envolve estabelecer relações entre os diversos tipos de recursos utilizados. Os citados abaixo são os mais importantes em termos organizacionais, tornando o material gerenciável e as informações acessíveis aos futuros usuários:

- Recursos empregados na documentação de uma língua.
- Recursos criados durante o trabalho de campo.
- Recursos que compartilham determinado gênero linguístico.
- Recursos referentes às diferentes mídias (vídeo, som, etc.) utilizadas na gravação.
- Recursos das transcrições e anotações referentes à determinada mídia de som.
- Recurso do léxico extraído das anotações.

As descrições fornecidas pelos metadados são responsáveis pelas informações de cada um desses recursos. Há dois tipos de metadados (citados em seção anterior), largamente utilizados na organização dos recursos empregados durante o processo de documentação linguística com diferentes finalidades, porém com funcionalidades complementares: o OLAC, projetado para facilitar a pesquisa em ambiente de metadados integrados cuja função assemelha-se ao catálogo de uma biblioteca; e o IMDI (baseado em esquema XML) que além de acumular as funções do OLAC, consegue gerenciar, hierarquicamente, o corpus documental. A figura abaixo indica o uso do metadados IMDI pelo DOBES para representar uma organização hierárquica de recursos.

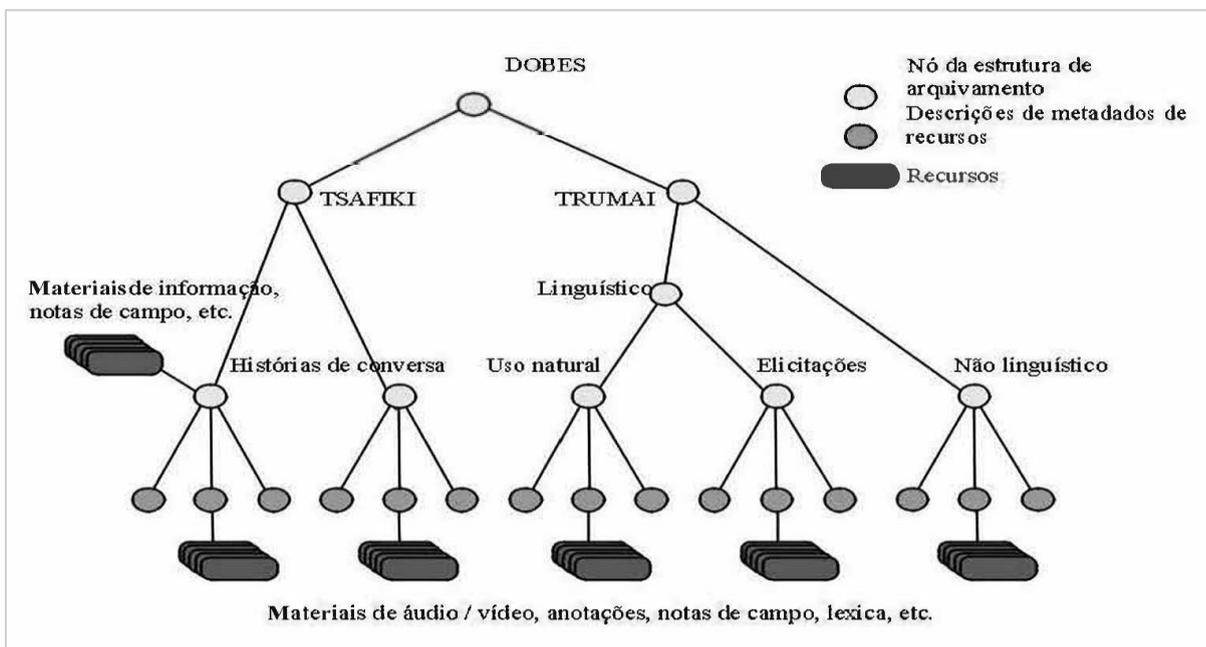


Figura VI: Esquema de corpus IMDI (DOBES) sobre a organização hierárquica de recursos<sup>83</sup>

4) a tarefa de armazenar é um ponto vital no processo de documentação linguística, por exigir o trabalho conjunto de arquivistas e depositantes no sentido de compatibilizar a obrigação do primeiro em receber o material enviado e guardá-lo, adequadamente, e, do segundo, em entregar o material nas condições ideais. Isso porque existem duas possibilidades de arquivamento: 1) em arquivo onde ocorre alguma perda na qualidade do material depositado independentemente de seu formato; 2) em arquivo com o formato aberto (não proprietário)<sup>84</sup> como padrão, exige-se que todos os objetos arquivados sigam esse formato. As duas realidades na relação com os arquivos impõem as seguintes possibilidades:

- A prevalência no uso dos arquivos de formato aberto por serem mais amigáveis aos usuários devido à facilidade de uso e, para os depositantes,

<sup>83</sup> Figura retirada e adaptada de Trilsbeek & Wittenburg (2006).

<sup>84</sup> Sobre os conceitos de software proprietário e não proprietário, podemos afirmar que: “Um formato aberto é uma especificação publicada para armazenar dados digitais, mantido geralmente por uma organização de padrão não-proprietário, e livre de limitações legais no uso, garantindo o acesso aos dados a longo prazo. Um formato aberto deve ser implementável tanto em software proprietário como em software livre, usando as licenças típicas de cada um. Em contraste, o formato proprietário é controlado e defendido por interesses particulares da empresa detentora de seus direitos, e caso algum software proprietário deixe de dar suporte àquele formato, desaparecendo do mercado, ele não poderá ser lido, logo, o usuário perderá informações contidas no mesmo, são exemplos: DWG (AutoCad Drawing), SWF (Shockwave Flash), DOC (formato proprietário de texto), PPT (formato proprietário de apresentação de slides). Os formatos abertos são um subconjunto do padrão aberto.” Disponível em: <[https://pt.wikipedia.org/wiki/Formato\\_aberto](https://pt.wikipedia.org/wiki/Formato_aberto)>. Acesso em: 01 nov. 2021.

pelos baixos custos para a conversão, manutenção da compatibilidade e preservação dos documentos digitais a longo prazo.

- A adaptação dos arquivos à recepção de formatos variados e a sua conversão caso seja necessário; e a solicitação aos depositantes que enviem documentos digitais em formatos compatíveis com a capacidade do arquivo, procedimento esse adotado pelos principais repositórios de línguas ameaçadas.

Independentemente de quaisquer das realidades expostas sobre os formatos aceitos pelos arquivos, é fundamental repetir a necessidade da exigência de que os depositantes enviem seus corpora digitais, com vistas ao arquivamento, fartamente documentadas por metadados, atitude que terá impacto direto no modo pelo qual as informações serão interpretadas futuramente por usuários, depositantes e arquivistas.

O acesso às informações linguísticas provenientes de todos os processos vistos até aqui é a última etapa das práticas da pesquisa documental, momento que ocorre após o depositante entregar o material digital ao arquivo. Para Austin (2015, p. 4), essa fase é aquela em que os dados são publicados – portanto, cuidados são necessários - geralmente para uso público, em ambiente World Wide Web (Internet) com perfis de acesso diferenciados para os usuários comuns, depositantes, especialistas de outras áreas, e comunidade de fala. Os privilégios diferenciados concedidos a cada um desses perfis certamente terão efeito na apresentação limitada ou não da corpora digital.

Bowern (2015, p. 225) se refere a essa etapa como sensível no que tange a diversos aspectos, pois entre outros questionamentos, a comunidade documentada pode discordar da ampla difusão de sua língua e cultura, razão pela qual essa possibilidade precisa ser discutida a partir de parâmetros éticos e legais na fase da pesquisa, como vimos em seção precedente.

Para os diferentes perfis de usuários, Trilsbeek & Wittenburg (2006, p. 325-327) definem as possíveis prioridades em relação ao conteúdo a ser consultado ou alterado (definido pelo arquivista e o depositante) no acesso ao arquivo, como:

- Usuários em geral: verificar anotações acompanhadas de léxico e/ou mídias de gravações, explicações culturais sobre a construção de uma casa acompanhadas de fotos, etc.

- Depositantes / pesquisadores: investigar, por metadados, a “codificação linguística contida no léxico em conjunto com notas etnográficas” ou adicionar novas anotações, etc.
- Membros da comunidade de fala: consultar para estudo, entretenimento, sugestão de correção, autoconhecimento, enriquecer o material, etc.

Por fim, podemos afirmar que a maioria dos arquivos responsáveis pelo armazenamento dos materiais linguísticos se orientam pelos parâmetros relativos à prática documental tratada genericamente nesta seção. Entretanto, deve-se considerar que o ALD não funciona isoladamente. Parte significativa deles é vinculada à Programas de documentação responsáveis pela sistematização de aspectos fundamentais quando o assunto é documentação linguística, como: projetos, recursos financeiros, métodos, teorias; enfim, o fornecimento da infraestrutura necessária para essa atividade. Em vista disso, alguns deles precisam ser analisados sob lupa perspicaz devido às particularidades que os caracterizam e os distinguem diante dos demais. Assim, pelos motivos citados associados aos limites teóricos de nossos estudos, resolvemos destacar em nossas investigações três desses programas: DOBES, DMA e AILLA.

## 2.5 – PROGRAMAS DE DOCUMENTAÇÃO

Conforme vimos, no bojo das previsões sombrias de Krauss (1992) e outros linguistas para o final do século XXI se previu que das 7000 línguas faladas no mundo, somente um terço continuaria a existir. A crítica situação transformou a questão das línguas ameaçadas em um novo campo de estudos e análises que explicam uma série de mobilizações acadêmicas cujo impacto, na área da Linguística, culminou em diversas ações, entre as quais o surgimento de uma nova subárea de pesquisa: a documentação linguística.

Como reflexo desse contexto surgiram, na passagem do século XX para o século XXI, diversas iniciativas voltadas ao arquivamento de corpora documental resultante de projetos de documentação linguística implementados em atividade de campo. Desde então, são vários os programas de documentação criados. Os financiados por fundações privadas sem fins lucrativos, os que são parte integrante de universidades ou centros de pesquisa. Alguns foram citados em seções anteriores desta dissertação.

Em nosso caso, optou-se por averiguar o modo como se estruturam três programas de documentação (DOBES, DMA e AILLA), o primeiro, vinculado à fundação privada sem fins lucrativos (Instituto Max Planck), e os demais, partes de duas instituições de ensino superior norte-americanas (Universidade de Chicago e Universidade do Texas). As análises serão feitas com base nas informações descritas em livros, artigos, nos códigos de conduta ou documentos similares expostos nos portais eletrônicos dos programas (pesquisa documental). Tais informações são divididas em duas diretrizes<sup>85</sup>: internas e externas. Sendo essa a maneira encontrada para diferenciar as instruções restritas ao âmbito interno dos programas daquelas que norteiam os princípios de suas relações com possíveis usuários que podem vir a se agregar (pesquisadores, consultores, comunidades, etc.); assim como nos orientar pelos princípios defendidos por Himmelman (2006,2007) em torno das melhores práticas na documentação linguística.

As diretrizes internas contêm informações que explicitam características organizacionais, financeiras e linguísticas dos arquivos digitais, como: objetivo, fonte

---

<sup>85</sup> Interprete-se, em nosso caso, o sentido mais apropriado para “diretriz”, como: “instrução ou orientação que deve ser seguida para levar a bom termo determinada tarefa; norma de procedimento”. Disponível em: <<https://www.infopedia.pt/dicionarios/lingua-portuguesa/diretriz>>. Acesso em: 25 nov. 2021.

de recursos, número de projetos, número de bolsas, método de documentação, ética no campo, método de arquivamento e política de acesso aos dados. Apesar desta última ser considerada uma diretriz interna, seus princípios objetivam a relação com o público externo.

As diretrizes externas contêm informações na forma de princípios normativos que explicitam a política de acesso ao arquivo linguístico digital (Código de Conduta) por intermédio de diversas cláusulas, como: 1) propriedade intelectual e privacidade dos indivíduos; 2) uso não comercial de dados; 3) e transferência de dados para arquivos externos. As diretrizes externas selecionadas são consideradas críticas dentre as demais por alguns pesquisadores que analisaram esse documento (WIDLOK, 2010; SOARES, 2007; DWYER, 2006). É importante ressaltar que em alguns casos os itens que formam o conteúdo de ambas as diretrizes estão diluídos ao longo dos tópicos referentes aos arquivos linguísticos digitais DOBES, DMA e AILLA.

A partir do mesmo enunciado, no final de cada subseção referente aos PDL (DOBES, DMA e AILLA), apresentaremos, em “Observações críticas”: 1) a resposta ao problema (teórico) desta dissertação (a partir de determinados princípios teórico-metodológicos) acompanhada da avaliação do programa de documentação; e 2) a análise de cada ALD (associado ao respectivo programa de documentação). O exame pormenorizado dos ALD obtiva obter subsídios para a proposta de projeto de arquivo linguístico para o Setor de Linguística do Museu Nacional.

## 2.5.1 – DOBES

### 2.5.1.1 – Antecedentes

Como vimos, os anos 1990 foram tensos para a comunidade ligada à Linguística no que se refere ao consenso acerca das graves denúncias sobre possibilidade de extinção de parte significativa das línguas ameaçadas até o final do século XXI. Diante desse quadro, instituições como a *Volkswagen Foundation*<sup>86</sup>

---

<sup>86</sup> A criação da Fundação Volkswagen (direito privado) tornou-se possível a partir da transferência de recursos feita pelo governo federal alemão e pelo estado federal da Saxônia após privatização da empresa montadora de veículos, no início da década de 1960, conhecida atualmente como Volkswagen AG. Não existindo, portanto, qualquer afiliação ou dependência entre as partes. Com uma reserva anual de 150 milhões de euros, tornou-se a maior financiadora privada de pesquisas na Alemanha voltada a instituições acadêmicas e desde 1962 apoiou mais de 33.000 projetos. Disponível em: <<https://www.volkswagenstiftung.de/en/foundation/about-us>>. Acesso em: 10 nov. 2021.

iniciaram, a partir de 2000, conforme suas diretrizes, atividades visando apoiar a documentação dessas línguas através do programa DOBES (captação de recursos e financiamento de projetos). Pelo que vimos, é inegável que tal iniciativa exerceu um grande protagonismo em diversos aspectos da documentação linguística desde o início de suas atividades, um mérito que também deve ser dividido com a iniciativa coletiva dos pesquisadores.

No portal DOBES<sup>87</sup> estão resumidamente expostas, as estruturas de seu funcionamento direcionadas à pesquisa documental, ancoradas em três eixos fundamentais relacionados a sua diretriz interna:

- A documentação linguística direcionada a três objetivos principais: 1) manutenção e revitalização linguística; 2) preservação do conhecimento sobre a diversidade linguística e tesouros culturais da humanidade para gerações futuras de falantes e pesquisadores; 3) introdução da quantificação na pesquisa linguística.
- A tecnologia avançada como ferramenta essencial para que seja possível a implementação de suas práticas documentais, principalmente aquelas voltadas ao arquivamento de longo prazo (formatos de gravação e análise, anotação de áudio e vídeo, criação de metadados e desenvolvimento de redes avançadas para o acesso a fontes de arquivos).
- Emprego de dois comitês de coordenação responsáveis pelo acompanhamento e orientação das atividades desempenhadas pelo DOBES: enquanto o Comitê Executivo responde pelas questões de interesse comum, o Comitê de Avaliação orienta arquivistas e pesquisadores no processo de documentação.

Na análise dos acervos linguísticos digitais feita até aqui, deve-se destacar a hegemonia do Programa DOBES em termos de recursos investidos, capilaridade geográfica dos projetos implementados e desenvolvimento de softwares voltados à análise linguística sobre o corpus coletado nas pesquisas de campo. O *The Language Archive* (TLA) é o arquivo central e se situa no Instituto Max Planck de Psicolinguística,

---

<sup>87</sup> Cf. DOBES Programme. Disponível em: <<https://dobes.mpi.nl/dobesprogramme/?lang=pt>>. Acesso em: 11 nov. 2021.

situado em Nijmegen, Holanda. O TLA (ramo tecnológico e organizacional do acervo DOBES) tem duas tarefas principais: 1) o arquivamento e preservação das gravações e dos tipos variados de dados linguísticos; e 2) o desenvolvimento de ferramentas direcionadas às atividades linguísticas, no nível tecnológico, voltadas ao arquivamento como um de seus objetivos finais, e, primordialmente, úteis para as atividades de campo.

#### 2.5.1.2 – Recursos

Até o último ano em que arrecadou fundos (2001-2012), foram mais de 60 projetos (ver Anexo F) em diversos locais no mundo, milhões de euros investidos no cenário em que seus métodos, teorias, orientações de melhores práticas, ferramentas e padrões técnicos seguem contribuindo para o desenvolvimento da documentação (ver anexos G e H). Conforme Franchetto & Rice (2014, p. 252-254), em continentes, como a América, e países como o Brasil, particularmente, o programa de financiamento DOBES teve reflexos profundos em pelo menos três aspectos:

- 1) Número de financiamentos realizados: os primeiros financiamentos no continente americano começaram em 2002 e os últimos terminaram em 2012. Foram cinco projetos na América do Norte, dois projetos na América Central e catorze projetos na América do Sul e a documentação de vinte e quatro línguas (ver Anexo I), 9 das quais no Brasil. Conforme Franchetto e Rice (2014, p. 254), no período, os 203 projetos executados em toda a região por DOBES, ELDP e ELF (maiores financiadores), dividiram-se assim: 70 (América do Sul), 55 (América Central) e 78 (América do Norte).
- 2) Surgimento de programas de financiamento nacionais voltados à documentação linguística: uma das consequências dos investimentos nos projetos de documentação no Brasil teve como resultado o surgimento de duas instituições nacionais voltadas à documentação das línguas indígenas: O Projeto de Documentação de Línguas Indígenas (PRODOCLIN), sediado no Museu do Índio, no Rio de Janeiro, em parceria com a Fundação Nacional do Índio (FUNAI), a Fundação Banco do Brasil e UNESCO, no ano de 2009 (ver lista de projetos de documentação DOBES sobre línguas indígenas

brasileiras<sup>88</sup>); e o arquivo linguístico localizado no Museu Emílio Goeldi, no Pará, desde a primeira década do século XXI. Ambos possuem metodologia documental própria voltada à realidade brasileira, mas a tecnologia de arquivamento utilizada pertence ao TLA que gerencia onde os materiais linguísticos coletados são armazenados. Nesse caso, não se sabe se localmente ou na matriz alemã (as informações presentes no portal do Museu Paraense Emílio Goeldi (MPEG) não são claras sobre a questão (ver Anexo J). Além dos acervos citados, existem outros dois na América do Sul que seguem os mesmos princípios: Laboratório de *Documentación y Investigación en Lingüística Y Antropología* (DILA), localizado na Argentina e o *Instituto de Investigaciones para la Amazonia peruana* (IIAP), situado no Peru (MOORE & GALUCIO, 2016; FRANCHETTO & RICE, 2014).

- 3) Crescimento da documentação linguística no Brasil: o crescimento dos financiamentos a projetos voltados à documentação linguística na América Latina por intermédio das agências de financiamento privadas estrangeiras (DOBES, ELPD, ELF, entre outras) - concentrado principalmente na primeira década do século XXI -, seguidas, posteriormente, por agências de fomento de caráter estatal, como as citadas anteriormente, teve dois objetivos claros: responder de modo pragmático à questão das línguas ameaçadas e acelerar o processo de documentação nos locais onde havia maior diversidade linguística na região (México, Guatemala e Brasil) com vistas a sua preservação e/ou revitalização. Nesse sentido, como afirmam Moore e Galucio (2016, p. 39): “um dos efeitos dos programas internacionais de documentação foi estimular a criação de um programa brasileiro para a documentação de idiomas indígenas usando a moderna tecnologia digital e a metodologia mais recente”.

---

<sup>88</sup> Ver DOBES (programas, projetos e instituições). Disponível em: <<http://prodoclin.museudoindio.gov.br/index.php/component/content/category/25-documentacao-linguistica>>. Acesso em: 27 nov. 2021.

### 2.5.1.3 – Arquivo linguístico digital

Os métodos de arquivamento do DOBES para línguas ameaçadas têm as seguintes características, como apontam Wittenburg, Mosel & Dwyer (2002) quanto ao material armazenado:

- Armazenamento a longo prazo.
- Controle personalizado de acesso aos dados.
- Acesso aos dados via Internet.
- Abertura dos metadados do material armazenado.
- Acesso e uso dos dados do arquivo adaptável a atuais e futuros usuários.
- Corpus linguístico multidisciplinar disponível a linguistas, etnólogos, educadores, jornalistas e, principalmente, povos indígenas.

Conforme as recomendações de Himmelmann (2007, 2006), o arquivo deve priorizar os interesses dos povos indígenas. Para isso, os pesquisadores são fundamentais nos pontos destacados acima no que se refere à elaboração do material exclusivo para as comunidades de fala. Nesse caso, o método de arquivamento é relevante, ao envolver normas que refletem a preocupação com o armazenamento de longo prazo, exigindo a utilização de mídias de gravação com esse escopo. Em linha com essas especificações, o arquivo DOBES utiliza as seguintes práticas organizacionais:

- Gravação de dados em diversos discos rígidos (HD), em diferentes locais geográficos.
- Transferência a cada cinco anos (tempo médio de duração de um HD) dos dados linguísticos para novos HD, automaticamente, por meio de algoritmos.
- Treinamento constante das atuais e futuras gerações de administradores de banco de dados para a preservação dos arquivos linguísticos indefinidamente.

Como vimos, conforme os princípios de melhores práticas estabelecidos pelo DOBES na documentação linguística, o pesquisador deve priorizar software não-proprietário (livre) no processo de documentação dos dados que serão depositados no arquivo. Os mesmos tipos de software, pelos fatores citados, deverão ser os responsáveis, no futuro, pela decodificação de todo arcabouço documental, na forma de metadados, que acompanha o corpus digital. Segundo Wittenburg et al. (2002), essa tarefa de interpretação de dados no DOBES ocorre em dois níveis: 1) o nível de “codificação linguística” ligado a aspectos tipológicos usados pelo DOBES; e 2) o nível “formal de codificação” que orienta a utilização de padrões abertos de codificação cujo detalhamento pode ser visto e analisado conforme o anexo, figuras VI e VII.

Além da preocupação com o armazenamento de longo prazo do corpus documental, o DOBES também priorizou, conforme suas diretrizes externas, a elaboração de políticas de acesso ao arquivo. Esse procedimento é controlado pela relação entre o usuário e os quatro níveis diferenciados de perfis<sup>89</sup> destacados a seguir:

- Nível 1: estão localizados os arquivos que podem ser abertos de modo bastante limitado por qualquer usuário (equivalente ao acesso aberto, público e gratuito).
- Nível 2: estão localizados os arquivos com abertura liberada a usuários registrados desde que este tenha assinado o código de conduta. A prioridade é a proteção dos dados conforme as exigências de direitos autorais. Nesse caso, são feitos acordos entre comunidades inteiras ou falantes individuais e o pesquisador (depositante) onde este tem a obrigação de controlar o acesso às fontes.
- Nível 3: estão localizados os arquivos acessíveis aos usuários após a aprovação do pesquisador (depositante) responsável, com garantia de acesso às fontes.
- Nível 4: estão localizados os arquivos quase totalmente fechados devido a aspectos delicados, como questões éticas, mas com a possibilidade de

---

<sup>89</sup> Cf. Acesso e registro DOBES. Disponível: <[https://dobes.mpi.nl/access\\_registration/?lang=pt](https://dobes.mpi.nl/access_registration/?lang=pt)>. Acesso em: 16 nov. 2021.

liberação futura após o pesquisador (depositante) especificar a data de sua abertura.

### 2.5.1.3.1 – Código de conduta

Considerando as diversas controvérsias que envolvem questões éticas durante o trabalho de campo, algumas das quais tratadas anteriormente, pontuamos a partir daqui o modo pelo qual o programa DOBES lida com o tema. Conforme Bower (2008) uma “pesquisa ética” pode ser definida como “uma maneira de trabalhar em que consultores, pesquisadores e a comunidade de fala, de comum acordo, julgam apropriada”, sem que exista qualquer tipo de normatividade institucional pré-estabelecida. Por outro lado, conforme as declarações gerais pertinentes ao Código de Conduta<sup>90</sup> DOBES, editado por Wittenburg (2005), estão colocadas as normas e posições éticas prescritas para pesquisadores, arquivistas, usuários e agências de financiamento, exigindo o respeito às regras, com a previsão de punições que preveem o afastamento do programa e o impedimento de acesso ao arquivo. A Figura VII apresenta a estrutura organizacional do Programa DOBES sujeita ao Código de Conduta:

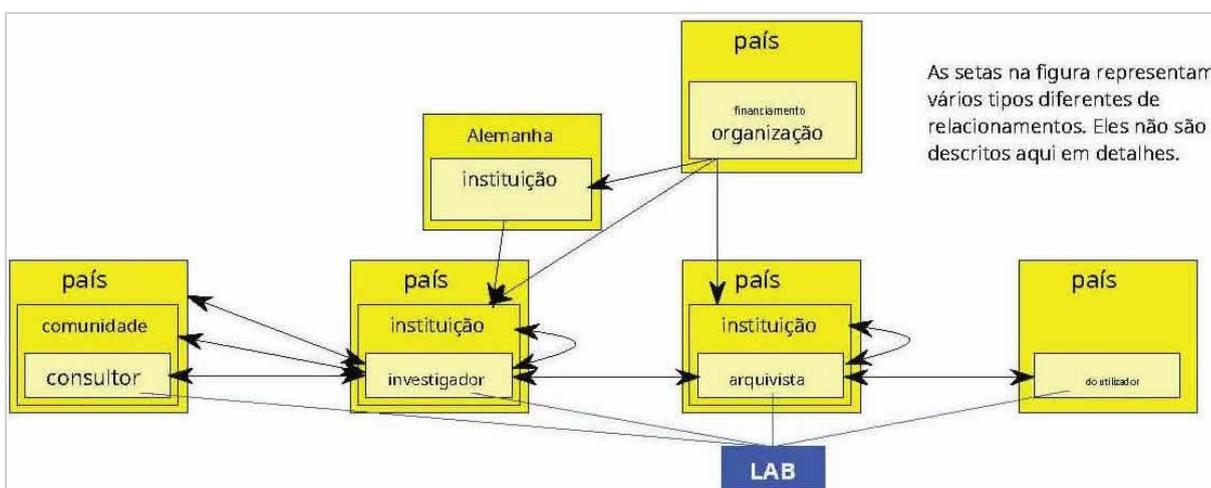


Figura VII: Abrangência do código de conduta DOBES sobre a estrutura organizacional<sup>91</sup>

<sup>90</sup> O Código de Conduta - e boas práticas - surge no final do século XX. Conforme Alemán (2007) apud Saddy (2017), os objetivos de sua criação estão, entre outros, em 'limpar a imagem de empresa ou profissional prejudicada por prática pouco ética, divulgar imagem positiva de empresa ou profissional, oferecer maior confiança ao consumidor ou usuário, funcionar como uma forma a mais de publicidade e de aumentar o êxito (ALEMÁN 2007 apud SADDY 2017).

<sup>91</sup> Cf. Code of conduct DOBES. Disponível: <[https://dobes.mpi.nl/ethical\\_legal\\_aspects/DOBES-coc-v2.pdf](https://dobes.mpi.nl/ethical_legal_aspects/DOBES-coc-v2.pdf)>. Acesso em: 16 nov. 2021.

As equipes DOBES (indicadas na figura na cor mais clara) compõem sua estrutura organizacional a partir de diferentes estruturas legais. Essas entidades se orientam por um código de conduta universal (para as ramificações do programa espalhadas pelo mundo) que abrange também os usuários dos arquivos. No caso particular dos arquivistas, conforme as regras estabelecidas, criou-se um Conselho consultivo linguístico responsável pelo ajustamento institucional às práticas internacionais de documentação das línguas ameaçadas.

O Código de Conduta (CdC) é composto por quinze tópicos<sup>92</sup> em que alguns se repetem em torno de algumas recomendações. Uma análise detalhada, em algumas delas, indica pontos “críticos” que, conforme o número de ocorrências, acreditamos serem dignas de atenção, como: respeito a direitos de propriedade intelectual (2 vezes), respeito às leis locais e nacionais (3 vezes), registro de consentimento formal (3 vezes), proibição sobre uso de dados para fins comerciais sem autorização (4 vezes), recomendações para arquivamentos e registros conforme padrões internacionais (3 vezes), compromisso do arquivista em garantir a integridade do corpus de dados no longo prazo (3 vezes), entre outros. A repetição das mesmas regras no CdC DOBES nos permite inferir que se tratam de questões complexas e sensíveis na relação do arquivo com os usuários.

Além das recomendações acima, de caráter normativo, oriundas do próprio programa, há aquelas feitas pelo judiciário internacional que norteiam os participantes do trabalho de campo em torno de sua relação com a comunidade de fala e do que se cogita extrair ao menos em três prerrogativas jurídicas que tratam também da submissão do programa à força das leis da UE em território holandês. De acordo com Dwyer (2006), no que tange à aplicação dos preceitos éticos no campo, pesquisadores costumam alternar entre soluções normativas e não normativas. Enquanto iniciativas individuais de pesquisa seguem a última opção, equipes mais completas (pesquisadores, usuários e arquivistas) optam pelas orientações presentes no código de conduta. Em linha com essa discussão, Himmelmann (2006, p. 16) afirma que: “[...] em meio a problemas e limitações éticas, mesmo os projetos de documentação mais cuidadosos têm o potencial de alterar profundamente a estrutura social da comunidade de fala documentada [...]”. Para mais discussões sobre ética e documentação linguística, ver Wilkins, 1992; Himmelmann, 1998; Grinevald, 2003.

---

<sup>92</sup> Cf. Regras éticas DOBES. Disponível em: <[https://dobes.mpi.nl/ethical\\_legal\\_aspects/DOBES-coc-v2.pdf](https://dobes.mpi.nl/ethical_legal_aspects/DOBES-coc-v2.pdf)>. Acesso em: 12 nov.2021.

Em contexto em que se aprofunda o emprego da documentação linguística digital, com destaque para o arquivamento e o controle de acesso aos dados no ALD que dependem das diretrizes definidas no Código de Conduta, identificamos dois aspectos na forma de acesso que têm se refletido diretamente em alguns conflitos de direitos envolvendo usuários: o padrão de acesso ao arquivo é aberto, público e gratuito no nível 1, mas com restrições crescentes conforme chega aos níveis 2, 3 e 4; o excesso de poderes do pesquisador (depositante) após o acesso ao arquivo. Quando é feita a relação entre esses dois aspectos, o ALD apresenta conflitos nas seguintes diretrizes externas: propriedade intelectual e privacidade dos indivíduos, uso não-comercial do corpus documental e transferência de dados para arquivos externos.

Reproduzimos adiante, literalmente, os tópicos das diretrizes externas (“Regras éticas”) que compõem o CdC DOBES destacados anteriormente. Os pontos críticos abaixo, cuja quantidade e temas refletem nossa proposta de análise, no primeiro momento podem indicar clareza na relação do usuário com os dados, porém a análise mais apurada suscita questionamentos:

- 1) “Todos respeitarão os Direitos de Propriedade Intelectual e Cultural dos consultores individuais e suas comunidades. Os desejos com relação à proteção da privacidade dos indivíduos serão respeitados.”
- 2) “Ninguém tem permissão para usar os dados gravados e analisados com propósitos comerciais sem permissão da comunidade de fala.”
- 3) “A transferência de dados DOBES para terceiros ou organizações são geralmente proibidas.”

#### 2.5.1.4 - Observações críticas

##### 1) Sobre o PDL DOBES

Conforme anunciado no início desta seção, as análises críticas dos programas de documentação investigados nesta dissertação, têm relação com seus esforços de responder à crise global de ameaça à diversidade linguística, denunciada nos anos 1990. A análise dessas medidas é feita a partir de três princípios estruturantes de um PDL definidos por Himmelmann (2007, p. 46–47), e consagrados por seus congêneres, como: 1) o enfrentamento à crise da diversidade linguística; 2) o

aproveitamento dos recursos financeiros destinados à pesquisa documental; e 3) a verificabilidade dos dados primários por pares. Abaixo, avaliamos de que modo o Programa DOBES implementou esses princípios:

- O problema da crise da diversidade linguística foi respondido com: a) as medidas de priorização de documentação daquelas com maior diversidade linguística, principalmente as situadas na América Latina; b) a aplicação de fortes recursos tecnológicos em todo o processo documental, principalmente na implementação de banco de dados linguístico; c) o envolvimento de universidades europeias e norte-americanas com novos projetos, etc.
- O aproveitamento de recursos financeiros ocorreu: a) mediante investimento em projetos de documentação em todos os continentes; b) concessão de bolsas de estudos; c) contínuo investimento em pesquisa, etc.
- A verificabilidade dos dados primários é possível após: a) a preparação da análise, tradução, transcrição e anotação dos metadados; b) a transferência destes para o arquivo linguístico; c) a publicação dos recursos (corpus) linguísticos no ALD. A discussão detalhada desse tema é realizada no tópico seguinte.

Como vimos, o Programa DOBES desde o início de suas atividades, há 25 anos, nunca foi unanimidade entre os linguistas. Por várias vezes foi criticado pelas tentativas de impor seus métodos (RODRIGUES, 2009) e melhores práticas (AUSTIN, 2012). Contudo, é inegável a sua grande contribuição para o avanço dos projetos e pesquisas em documentação linguística.

Por fim, pelas discussões feitas no início desta seção, complementadas por essas observações, respondemos ao problema teórico desta dissertação, avaliando que o Programa DOBES cumpre, plenamente, aos princípios estruturantes de um PDL ao combater a crise global da diversidade linguística, em geral, e, particularmente, a ameaça às línguas indígenas.

## 2) Sobre o ALD DOBES

A partir da relação entre a política de acesso com os pontos críticos do CdC assinaladas (WIDLOK 2015, p. 8-10; SOARES, 2007, p. 2), podemos concluir que:

- O arquivo no modo aberto, público e gratuito tem o uso bastante limitado (geralmente permite somente visualizar metadados). Porém, o padrão de acesso ao arquivo é fechado para a maioria dos acessos (veja, acima, os parâmetros para o usuário acessar o arquivo nos níveis 2, 3 e 4), principalmente devido à pressão da comunidade (pesquisadores, consultores e indígenas) mesmo que personalidades acadêmicas importantes (HIMMELMANN 2007, 2006) defendam o arquivo aberto.
- O arquivista e o técnico transferem para o pesquisador (depositante) a responsabilidade pela liberação de acesso ao arquivo para o usuário, situação que ocorre no nível 3, geralmente, para estudantes de doutorado, mesmo que as atividades de campo não tenham sido finalizadas e a formação do corpus documental não tenha sido concluída. Acontecem, nesse momento, conflitos de direitos de propriedade intelectual e privacidade que ferem o Código de Conduta.
- O acesso privilegiado ao arquivo DOBES, concedido ao pesquisador (depositante) permite, como vimos, que este libere, a elementos exógenos ao grupo de pesquisa titular, o acesso a materiais sensíveis e inéditos ainda na fase de desenvolvimento. Esta situação rompe, a princípio, com os dois primeiros direitos destacados no CdC (propriedade intelectual e privacidade) e eleva, sobremaneira, a possibilidade de que os dois últimos (uso não comercial e transferência de dados para arquivo externo) sofram algum tipo de violação.
- A possibilidade de que relações interpessoais influam na definição de níveis de acesso com privilégios impróprios para determinados usuários pode afetar alguns preceitos do CdC. Entretanto, pode ser um sinal de que é necessário o estabelecimento de vínculos interinstitucionais que permitam o acesso e compartilhamento de dados dos arquivos linguísticos entre organizações que possuam ALD.

A título de esclarecimento, note-se que a “transferência de dados para o arquivo

externo” não pode ser confundida com o conceito de “mobilidade”, um dos fundamentos da documentação linguística explicitada por Himmelmann (2007, 2006) que corresponde à possibilidade técnica de publicação do corpus digital na Internet.

Por fim, podemos afirmar que as limitações impostas para o acesso ao arquivo no modo aberto, público e gratuito (nível 1) força os usuários ao cadastro nos níveis mais altos (2, 3 e 4) que os submetem, respectivamente, aos seguintes crivos: registro e assinatura do Código de Conduta (níveis 2, 3, 4), aprovação do depositante (nível 3) e liberação de arquivo fechado somente pelo depositante (nível 4). Como visto, além de ser obrigado a se cadastrar, o usuário interessado no corpus documental armazenado no ALD precisa se resignar diante dos “superpoderes” atribuídos ao depositante.

A cadeia de interesses que, implicitamente, força cadastramentos por um lado, e por outro, controla o acesso dos cadastrados (negando o acesso a certos arquivos para uns e liberando o acesso dos mesmos para outros), ao mesmo tempo que confirma a importância do cumprimento das regras éticas que garantem os direitos de pesquisadores, consultor(es) e comunidade de fala, ressalta comportamentos obscuros por parte de arquivistas e técnicos que se omitem na sombra do depositante, cujo controle sobre o ALD quase o transforma no “gerente geral” único de um patrimônio de informações que desperta o interesse da coletividade envolvida com o problema das línguas minoritárias e, particularmente, das línguas indígenas ameaçadas de extinção (WIDLÖK 2015).

Em relação às principais críticas feitas até aqui, ressalte-se dois pontos importantes: 1) o controle dos usuários através dos níveis de acesso é crucial para o funcionamento de qualquer ALD. Porém, é necessário que após o cadastro do usuário, conforme o nível, esse acesso permita o trabalho com o corpus documental da forma mais satisfatória possível no processo de pesquisa. 2) Acerca do maior poder do arquivista sobre a atribuição de privilégios de acesso, essa prerrogativa tem relação direta com seu vínculo institucional de empregado vinculado a empregador, empresa ou órgão público, com a existência de chefia imediata à qual é subordinado, ou seja, nem sempre ele poderá tomar decisões isoladas. Enquanto o depositante é alguém que é raramente conhecido pessoalmente, por utilizar o ALD virtualmente, situação em que, por vezes, o máximo de informações que se tem dele está no formulário eletrônico preenchido na Internet. Assim, é possível que as recomendações feitas talvez não resolvam completamente as possibilidades de ameaças à integridade do

arquivo, mas certamente tornarão o controle sobre elas mais efetivo.

Deste modo, é possível afirmar que o “desequilíbrio de poderes” entre os usuários do ALD impacta sobre os direitos de propriedade intelectual. Esse quadro, somado aos problemas de segurança característicos do ambiente virtual, representa maiores dificuldades diante de todo o aparato de filtros de segurança responsáveis pela preservação da integridade dos materiais arquivados. Deste modo, não se pode garantir que o acervo esteja livre de ameaças a seu conteúdo, pois como afirmam Trilsbeek & Wittenburg (2006, p. 3): “[...] nenhum sistema de acesso pode impor totalmente o uso adequado de materiais de arquivo. Somente o controle social sobre a comunidade de usuários pode impedir que o material esteja sendo usado para outros fins que não os solicitados.”

## 2.5.2 – DMA

### 2.5.2.1 – Antecedentes

A história da Universidade de Chicago (UC)<sup>93</sup> é marcada, nas primeiras décadas do século XX, pela criação do primeiro laboratório, o *Scientific Study of Language Phenomena* por Clarence A. Parmenter. Professor do Departamento de Linguística e Línguas Românicas, de 1914 a 1953, tornou-se pioneiro nos estudos de teoria e metodologia fonética. Durante o período de tensões geopolíticas provocadas pela II Guerra Mundial e depois pela Guerra Fria, a instituição foi recrutada pelas forças armadas estadunidenses para ensinar línguas estrangeiras a civis e militares em zonas de conflito, atividade que envolveu a cooperação do linguista Norman A. McQown. Após o “esforço de guerra”, o professor dedicou-se à criação do Laboratório de Linguística que, com poucos recursos e instalações incompletas, funcionou de 1955 a 1962. Posteriormente, por intermédio de doação da Fundação Ford à Universidade de Chicago, um novo laboratório foi construído e voltou-se para os *for Non-Western Area Programs and Other International Studies*. Inaugurado em 1962 como *Ford Foundation International Studies Language Laboratory*, teve o nome alterado pela professora Karen L. Landahl, para *Language Laboratories and Archives* (LLA) em 1988.

---

<sup>93</sup> Cf. Digital Media Archive. Disponível em: <<https://dma.uchicago.edu/about/history>>. Acesso em: 06 dez. 2021.

Atualmente, com outros setores, a *Division of the Humanities*<sup>94</sup> da UC abriga o *Department of Linguistics* e o *Digital Media Archive (DMA)*. Com o fim do *Language Laboratories and Archives*, em 2006, todo o seu acervo foi transferido para o DMA, tornando-se este o arquivo digital responsável pelo armazenamento de vasto e diversificado material audiovisual, como: 1) o acervo físico (equipamento de áudio/vídeo, computadores, documentos, etc.) pertencente ao setor de processamento, arquivamento e preservação de mídias analógicas; e 2) materiais de gravação mecânicos (cilindros, fitas de áudio, discos fonográficos, fitas de vídeo, etc.), resultado de pesquisas de três gerações de pesquisadores ligados à Linguística, ensino de idiomas e à Antropologia. É importante ressaltar que o Departamento de Linguística da UC (fundado em 1930, é o mais antigo dos Estados Unidos) e, juntamente com outros setores da instituição, também foi atingido pela ampla preocupação com as línguas ameaçadas de extinção, desenvolvendo projetos e pesquisas em documentação linguística apoiados por agências de financiamento; tema que será analisado mais adiante.

O Programa DMA, criado em 2006, possui um acervo online cujo conteúdo é fornecido basicamente pelas pesquisas de documentação linguística contemporâneas e pelo material analógico audiovisual recebido do antigo LLA, encerrado no mesmo ano, o qual é gradualmente digitalizado e disponibilizado no portal do DMA<sup>95</sup>. Ou seja, diferentemente do que ocorreu com outras instituições que se dedicaram a investir no financiamento das pesquisas em documentação linguística (*Volkswagen Foundation*, *ELDP*, *ELF*, etc.) motivadas pela situação denunciada no início dos anos 1990 (KRAUSS, 1992), a UC já possuía, desde a década de 1950, um embrião de espaço dedicado à preservação dos estudos linguísticos multidisciplinares, por certo considerando as particularidades daquele contexto histórico e as especificidades dos estudos feitos.

---

<sup>94</sup> Cf. *Division of the Humanities*. Disponível em: <<https://humanities.uchicago.edu/>>. Acesso em: 06 dez. 2021.

<sup>95</sup> Cf. *Coleções do DMA*. Disponível em: <<https://dma.uchicago.edu/locations>>. Acesso em: 09 dez. 2021.

### 2.5.2.2 – Recursos

Em termos de financiamento para as diversas atividades de pesquisas, foram identificadas, na Universidade de Chicago, cinco setores acadêmicos que capitam recursos regularmente, com destaque para a documentação linguística, entre os quais destacam-se:

- O Programa DMA conta com um conjunto significativo de antigos acervos destinados, na maioria, à digitalização. Esse material é resultado de doações de linguistas, antropólogos, curso de ensino de idioma, pesquisas de documentação linguística, etc. No portal do programa<sup>96</sup> estão registradas distintas origens de recursos cujos montantes em valores absolutos não são revelados, apesar de ser possível verificar que vêm de pessoas físicas e jurídicas que aportam recursos financeiros ou liberam bolsas de estudo, com o destaque para o NEH, *Ford Foundation*, *Center for Latin American Studies*, *Linguistic Interviews* (John Goldsmith), *Carnegie Foundation*, entre outros. Há também outros tipos de doação, como as direcionadas à distribuição de material analógico de áudio e vídeo e documentos de algumas áreas do conhecimento (Linguística, Antropologia, Sociologia, Jornalismo, etc.) para a digitalização e publicação no arquivo DMA.
- O *Committee on Latin America Studies* (CLAS), criado em 1968 para coordenar os interesses universitários em pesquisa e ensino na América Latina, funcionando, atualmente, como local de intercâmbio intelectual com estudos voltados para a região<sup>97</sup>. As pesquisas são fomentadas por diversas contribuições (bolsas e financiamento), com destaque para as feitas pela NSF e a NEH, direcionadas ao programa *Documenting Endangered Languages* com o propósito de digitalização e catalogação de manuscritos etnográficos e linguísticos produzidos no início do século XX por instituições que trabalham com línguas e culturas indígenas da América Central.

---

<sup>96</sup> Cf. “About the contributors”. Disponível em: <<https://dma.uchicago.edu/about/about-contributors>>. Acesso em: 20 dez. 2021.

<sup>97</sup> Cf. Center for Latin American Studies (CLAS). Disponível em: <<https://clas.uchicago.edu/about/history>>. Acesso em: 12 jan. 2022.

- O *Karen Landahl Center for Linguistics Research* cujo objetivo é permitir aos professores, estudantes e professores associados aprofundar os conhecimentos sobre a natureza da linguagem e funciona no Centro de Pesquisa de Ciências Sociais. Abriga quatro laboratórios de pesquisa e ensino e dois grupos de pesquisa, entre os quais destaca-se o *Language in Time Space Research Group*<sup>98</sup>. Dirigido pela linguista Lenore Grenoble, tem por objetivo fornecer um espaço para pesquisa sobre línguas subdescritas em ambientes de trabalho de campo, com envolvimento ativo em todos os aspectos da documentação linguística, concentrando-se no conjunto de abordagens e metodologias próprios.
- O programa *Preserving Indigenous Mesoamerican Language Heritage*, financiado pelo NSF e coordenado pelo professor emérito John Lucy (especialista na relação entre linguagem e pensamento e na língua Yucatec Maya), tem como meta melhorar a infraestrutura das coleções de materiais da UC relacionadas à documentação das línguas indígenas mesoamericanas cujo acervo teve origem na década de 1920<sup>99</sup>. O arquivo é composto por gravações e textos sobre dezenas de línguas indígenas extintas ou ameaçadas de extinção. O projeto se propõe a desenvolver estruturas de metadados padrão para as coleções cujo conteúdo deve ser disponibilizado em portal eletrônico na Internet, tornando-se acessível a acadêmicos, comunidade de falantes e demais interessados.
- O programa de bolsas *Documenting Endangered Languages - Doctoral Dissertation Research Improvement Grants* (DEL DDRIG) financiado pela NSF<sup>100</sup>. Por meio deste, fomenta-se a utilização da tecnologia da informação com vistas à construção de infraestrutura computacional de modo a apoiar as pesquisas relacionadas às línguas ameaçadas de extinção por intermédio de gestão, arquivamento de dados e suporte ao

---

<sup>98</sup> Cf. Karen Landahl Center for Linguistics Research. Disponível em:

<<https://linguistics.uchicago.edu/research/research-groups-and-labs>>. Acesso em: 13 jan. 2022.

<sup>99</sup> Cf. John Lucy awarded NSF grant for preservation of mesoamerican language materials. Disponível em: <<https://clas.uchicago.edu/content/john-lucy-awarded-nsf-grant-preservation-mesoamerican-language-materials>>. Acesso em: 14 jan. 2022.

<sup>100</sup> Cf. Documenting Endangered Languages - Doctoral Dissertation Research Improvement Grants (DEL DDRIG). Disponível em: <<https://grad.uchicago.edu/fellowship/national-science-foundation-nsf-documenting-endangered-languages-doctoral-dissertation-research-improve%E2%80%A6>>. Acesso em: 15 jan. 2022.

trabalho de campo. Envolve a gravação digital, documentação e arquivamento de corpora das línguas ameaçadas, constituído por léxicos gramaticais, amostra de textos e banco de dados. Os valores dos financiamentos oscilam entre US\$ 150.000 e US\$ 250.000, sendo destinados, exclusivamente, a instituições universitárias norte-americanas.

Ao tratar de pesquisas envolvendo documentação linguística feitas, exclusivamente, no Brasil, conforme os referenciais definidos e já descritos, podemos destacar duas produções acadêmicas: 1) a dissertação de Adam Roth Singerman (“The Morphosyntax of Tuparí, a Tupían Language of the Brazilian Amazon”), defendida em 2018 sob orientação de Lenore Grenoble voltada à documentação e análise de línguas da Amazônia - nesse caso, o Tuparí em Rondônia -, com ênfase em fenômenos morfossintáticos de interesse tipológico e teórico<sup>101</sup> cujo período de pesquisas (2016-2018) teve como fonte de financiamento o apoio de diversas instituições, entre as quais ELF, NSF, CLAS, etc.; 2) e a tese de doutorado de Eduardo Rivail Ribeiro (“A Grammar of Karajá”), defendida em 2012 sob orientação de Amy Dahlstrom, na UC, que certamente tem relação direta com a documentação linguística, pois como ele mesmo afirma, trata-se de “pesquisa de campo e de arquivo” que reflete o aprendizado adquirido em projeto anterior voltado para línguas indígenas, com o título “Análise, descrição e documentação da língua Ofayé”<sup>102</sup>, elaborada entre 2002 e 2006 por intermédio de financiamento de programas internacionais de documentação para línguas indígenas no Brasil, em associação entre o ELDP (financiador), a Universidade de Chicago e a Universidade Federal de Goiás (UFG) - ver Anexo I.

O Departamento de Linguística da UC, inserido na estrutura da *Division of the Humanities Linguistics* conta, atualmente, com um corpo docente constituído por vinte e seis professores, dezoito dos quais são pertencentes ao *core faculty*, entre os quais destacam-se Lenore A. Grenoble e Salikoko S. Mufwene, os quais com algumas produções acadêmicas e pesquisas apoiadas por agências de financiamento voltadas às línguas ameaçadas de extinção no mundo sendo, portanto, comprometidos com a

---

<sup>101</sup> Cf. Adam R. Singerman. Disponível em: <<https://adamsingerman.wordpress.com/>>. Acesso em: 18 jan. 2022.

<sup>102</sup> Cf. Sistema de Currículos Lattes. Disponível em: <http://lattes.cnpq.br/8204770305516420>. Acesso em: 19 jan. 2022.

iniciativa de documentá-las.

No caso de Grenoble, além da participação em grupos de pesquisa, a pesquisadora possui diversos artigos e livros onde expressa sua preocupação e de seus pares com os rumos que vêm tomando a documentação linguística em seu processo de expansão. É isso, por exemplo, que indica a obra “Language Documentation: practice and values” (GRENOBLE & FURBEE, 2010). Resultado do trabalho de vários especialistas em documentação linguística, o livro teve o financiamento da *Linguistic Society of America (LSA)*, da NSF e da Divisão de Humanidades da Universidade de Chicago. O produto final é resultado dos consensos gerados a partir da Conferência das Línguas Ameaçadas, organizada pelo *LSA Summer Institute* em 2005, onde se discutiram os seguintes tópicos: “1) os requisitos da formação linguística de campo; 2) as preocupações e o envolvimento das comunidades linguísticas sobre seu patrimônio; 3) a questão do que é a documentação adequada; 4) os usos da documentação nas comunidades de fala; 5) treinamento e carreiras em linguística de campo; e 6) ética e melhores práticas de arquivamento”<sup>103</sup>.

Além da experiência de ter suas obras acadêmicas publicadas com o auxílio de agências de financiamento, as suas pesquisas de campo também acompanham essa tendência. Com destaque para a concessão da NEH e da NSF, em 2014, de um total de US\$ 4 milhões para pesquisas voltadas à documentação de 40 línguas. Nesse caso, a pesquisadora e professora da Universidade de Chicago, com o projeto intitulado “Understanding spatial determinants, complex predicates, and case marking through traditional narratives in endangered languages” foi contemplada com quase US\$ 240.000 através da NSF.

---

<sup>103</sup> Ibid., p. 3.

### 2.5.2.3 – Arquivo linguístico digital

Como vimos, a América Latina é objeto de estudos dos linguistas da universidade por intermédio de duas iniciativas diretas: o CLAS e o programa *Preserving Indigenous Mesoamerican Language Heritage*. Ambos, apesar de estarem envolvidos com pesquisas sobre documentação linguística voltadas às línguas ameaçadas de extinção, não especificam, em números, a dimensão e os locais de seus projetos. O contrário acontece em relação à amplitude dos dados fornecidos pelo Programa DMA. Em seu portal eletrônico, constam, aproximadamente, 2100 coleções linguísticas em 270 línguas e projetos desenvolvidos em todos os continentes<sup>104</sup>, efetuados em décadas anteriores e na atualidade, muitos dos quais em processo de digitalização, e, posterior, disponibilização.

Entre o conjunto de mídias disponíveis, há aquelas que são gravações analógicas (acumuladas por décadas) de áudio e vídeo de línguas indígenas (não possuem metadados, análises, anotações, etc.). Não sendo, portanto, consideradas estrita documentação linguística, porém, podem ser relevante manancial de estudos para projetos a serem desenvolvidos. Além disso, essas variadas coleções incluem notas técnicas e descritivas de gravações de áudio e/ou vídeo que englobam o ensino de língua estrangeira, música indígena, música popular, poesia, entrevistas com indígenas (no caso da coleção brasileira, ver Anexo K e L), etc. Na busca feita no ALD DMA, na Internet, para a localização de conteúdo referente a corpus documental no Brasil, são disponibilizados, como mencionado, diversos materiais, dentre os quais os textuais, mídias de áudio e vídeo, entre outros. Quanto às línguas indígenas cujos dados primários estão armazenados no ALD, estão presentes as seguintes etnias, metadados e abordagens, algumas delas acompanhadas de notas descritivas e técnicas<sup>105</sup>:

- Krenak e Kamaiura: “Interviews in Krenak (Borun) and Kamaiuré”.
- Kaiapó: “Kayapó sound tracks (Turner, c. 1979)”.

<sup>104</sup> Cf. Diferentes tipos de materiais arquivados no DMA referentes a várias regiões, estados, províncias e países do mundo, inclusive o Brasil. Disponível em: <<https://dma.uchicago.edu/locations>>. Acesso em: 15 jan. 2022.

<sup>105</sup> Cf. Digital Media Archive. Disponível em: <[https://dma.uchicago.edu/search?f%5B0%5D=dma\\_collection\\_locations%3A7663](https://dma.uchicago.edu/search?f%5B0%5D=dma_collection_locations%3A7663)>. Acesso em: 29 dez. 2021.

- Krenak: “Two Krenakôre myths (Schwartzman, 1981)”.
- Tenharim: “Psychological interviews in Tenharim (Parintintin) (W. Kracke, 1973)”.
- Tenharim: “Parintintin myths (W. Kracke, 1973); Tenharim myths”.

Diferentemente do que ocorre com o Programa DOBES, que possui diretrizes sobre sua proposta metodológica de documentação linguística, amplamente divulgadas no site, na Internet, e em diversas publicações acadêmicas, muitas das quais vistas anteriormente, pude constatar, por meio de busca intensa em portais especializados de publicações acadêmicas (SCIELO, Periódicos CAPES, Google Science, Biblioteca Digital de Teses e Dissertações, Science.Gov, WorldCat, etc.) que o Programa DMA não disponibiliza os parâmetros gerais das metodologias de pesquisa, de modo explícito, por meio de publicações tradicionais ou virtuais. Uma possibilidade talvez seja imaginar que os linguistas estejam mais suscetíveis às demandas de adequação às “melhores práticas” impostas pelas agências de financiamento do que, propriamente, a qualquer limitação metodológica exigida pela instituição acadêmica a qual estejam vinculados.

No caso específico do Programa DMA, existem padrões de formatos de arquivos e regras para o depósito dos materiais<sup>106</sup>. Podem ser arquivados materiais digitais ou analógicos após avaliação técnica. Exige-se que todos os depósitos sejam acompanhados de informações complementares para a adição de novos conteúdos ao arquivo, como: metadados, informações sobre direitos autorais, transcrições, notas de campo, etc. As gravações fora dos padrões exigência são avaliadas pelo Diretor do DMA em consulta ao Comitê Diretor, situação em que podem ser rejeitadas. Restringe-se também o recebimento de materiais digitalizados que não sigam os parâmetros de formatos exigidos para gravações de áudio e/ou vídeo, documentos (formulários e transcrições), bancos de dados e planilhas, prática semelhante à adotada pelos arquivos de grande porte.

A partir das informações existentes no site do Programa DMA, não está claro se exigências de formato de arquivo se limitam à documentação linguística, parece que não. Assim, pode-se concluir que as aquisições dos materiais, conforme as regras

---

<sup>106</sup> Cf. “Acquisitions guidelines”. Disponível em: <<https://dma.uchicago.edu/about/acquisitions-guidelines>>. Acesso em: 22 dez. 2021.

estabelecidas, são abertas a outras áreas da ciência. As exigências feitas aos depositantes, em termos de informações dos materiais digitalizados na forma de metadados, seguem os mesmos padrões estabelecidos por outros programas de documentação.

#### 2.5.2.3.1 – Código de conduta

Em termos complementares às regras e padrões voltados às políticas de aquisição e método de arquivamento, é importante salientar que o acesso ao arquivo DMA é aberto para as coleções públicas e gratuitas e sem exigência de protocolos de acesso. Por outro lado, o acesso às coleções não públicas (com restrições de acesso) é autorizado somente após o preenchimento de formulário específico e a autorização dos gestores da plataforma digital. Restrições de acesso ocorrem também nos casos de problemas com direitos autorais, não digitalização da mídia analógica e bloqueio no uso de doações determinados pelas concedentes enquanto estiverem vivas<sup>107</sup>. Para pesquisas mais detalhadas nas coleções exige-se, previamente, o preenchimento de formulário específico com informações necessárias ao cadastro (nome e sobrenome, endereço de e-mail, e descrição resumida do projeto de pesquisa com a lista das coleções para as quais deseja ter acesso) e a posterior aprovação ou não da solicitação<sup>108</sup>.

Como sabemos, o arquivo DMA é ambivalente no sentido de abrigar estrutura organizacional e tecnológica que permite o armazenamento de materiais físicos (analógicos ou não) e materiais digitais. Tais capacidades são reguladas por “Diretrizes de aquisição”<sup>109</sup> e “Política de aquisições”<sup>110</sup> direcionadas à digitalização, arquivamento e acessibilidade ao material depositado na Internet. Podemos considerar que essas recomendações reunidas seriam equivalentes, guardados os devidos cuidados, a um Código de Conduta, por explicitarem em seu bojo, mesmo que superficialmente, temas tecnicamente fundamentais que os assemelham. Como

---

<sup>107</sup> Cf. “About the Digital Media Archive”. Disponível em: <<https://dma.uchicago.edu/about/about-digital-media-archive>>. Acesso em: 22 dez. 2021.

<sup>108</sup> Cf. “Request Research Access”. Disponível em: <<https://dma.uchicago.edu/forms/request-research-access>>. Acesso em: 22 dez. 2021.

<sup>109</sup> Cf. “Acquisitions guidelines”. Disponível em: <<https://dma.uchicago.edu/about/acquisitions-guidelines>>. Acesso em: 20 fev. 2022.

<sup>110</sup> Cf. “Acquisitions policy”. Disponível em: <<https://dma.uchicago.edu/about/acquisitions-policy>>. Acesso em: 20 fev. 2022.

exemplo, podemos citar a relação do DMA com as diretrizes externas (propriedade intelectual e privacidade dos indivíduos, uso não-comercial do corpus documental e transferência de dados para arquivos externos) destacadas para analisar os pontos críticos na relação dos usuários com o arquivo digital que afetam alguns direitos da comunidade de usuários.

#### 2.5.2.4 – Observações críticas

##### 1) Sobre o PDL DMA

Conforme anunciado no início desta seção, as análises críticas dos programas de documentação investigados nesta dissertação, têm relação com seus esforços de responder à crise global de ameaça à diversidade linguística, denunciada nos anos 1990. A análise dessas medidas é feita a partir de três princípios estruturantes de um PDL definidos por Himmelmann (2007, p. 46–47), e consagrados por seus congêneres, como: 1) o enfrentamento à crise da diversidade linguística; 2) o aproveitamento dos recursos financeiros destinados à pesquisa documental; e 3) a verificabilidade dos dados primários por pares. Abaixo, avaliamos de que modo o Programa DMA implementou esses princípios:

- O problema da crise da diversidade linguística foi respondido com: a) o recebimento de doações de acervos analógicos multidisciplinares; b) digitalização de acervos analógicos e armazenamento no arquivo; c) desenvolvimento de projetos de documentação linguística na América Latina e Brasil.
- O aproveitamento de recursos financeiros ocorre: a) mediante investimento em projetos de documentação em todos os continentes; b) investimento em pesquisa por meio de estudantes bolsistas; c) investimento em equipamentos tecnológicos para digitalização dos acervos analógicos recebidos por doação, etc.
- A verificabilidade dos dados primários é possível após: a) a preparação da análise, tradução, transcrição e anotação dos metadados; b) a transferência destes para o arquivo linguístico; c) a publicação dos

recursos (corpus) linguísticos no ALD. A discussão detalhada desse tema é realizada no tópico seguinte.

Um dos principais problemas encontrados no PDL DMA, e, por conseguinte, no respectivo ALD, reside na absoluta falta de referências textuais sobre sua história, e seus objetivos com o programa de documentação e o arquivo linguístico. As investigações indicam falta transparência em diversos aspectos no PDL. Contudo, isso não invalida a iniciativa em si, pois o programa está ativo, captando recursos e investindo em novos projetos, inclusive no Brasil. Até momento do encerramento desta dissertação, o requisito “verificabilidade dos dados primários” não havia sido preenchido.

Por fim, pelas discussões feitas no início desta seção, complementadas por essas observações, respondemos ao problema teórico desta dissertação, avaliando que o Programa DMA cumpre, com ressalvas, aos princípios estruturantes de um PDL ao combater a crise global da diversidade linguística, em geral, e, particularmente, a ameaça às línguas indígenas.

## 2) Sobre o ALD DMA

Considerando a falta de um “CdC DMA” equivalente ao visto no DOBES (do ponto de vista do volume e da transparência das informações disponibilizadas), mesmo com a compilação das “Diretrizes de aquisição”<sup>111</sup> e da “Política de aquisições”<sup>112</sup>, percebemos que algumas questões de direitos mais sensíveis, indicadas nas diretrizes internas e externas são ignoradas. De modo que algumas omissões de políticas adotadas pelo ALD, principalmente relacionadas a direitos que poderiam implicar em prejuízos para a comunidade de usuários, nem são citadas. Assim, vamos nos ater a ressaltar aquilo que é dito e refletir sobre as consequências do que está oculto, ressaltando os seguintes pontos analisados a partir das Diretrizes e Política de aquisições:

---

<sup>111</sup> Cf. “Acquisitions guidelines”. Disponível em: <<https://dma.uchicago.edu/about/acquisitions-guidelines>>. Acesso em: 20 fev. 2022.

<sup>112</sup> Cf. “Acquisitions policy”. Disponível em: <<https://dma.uchicago.edu/about/acquisitions-policy>>. Acesso em: 20 fev. 2022.

- O Programa DMA tem como prioridade em suas diretrizes de aquisição a “Política de digitalização” direcionada à digitalização dos arquivos para preservar, arquivar e tornar acessíveis, no longo prazo, as coleções doadas. Conforme etapas previstas, os materiais adquiridos são depositados nos arquivos da universidade, digitalizados, e, posteriormente, armazenados no arquivo digital, segundo as permissões previamente acordadas de propriedade intelectual e direitos éticos.
- O padrão de acesso ao arquivo é aberto para as coleções públicas e gratuitas, mas é fechado para as coleções não públicas para as quais exigem-se protocolos de acesso (login e senha). São qualificados como depositantes apenas aqueles que possuem algum vínculo, direto ou indireto, com a universidade (professor, estudante de pós-graduação, pessoa ligada a algum membro do corpo docente ou que possui material relacionado a projeto de pesquisa vinculado ao programa).
- O depositante pode, autonomamente, confirmar os direitos de propriedade intelectual do conteúdo a ser arquivado e dos direitos éticos dos falantes e comunidade de falantes; assim como também pode tomar essa decisão após consulta ao Comitê do Programa DMA quando for depositar.
- A capacidade de uso do ALD pelo depositante é limitada. A explicação para isso reside em duas constatações: 1) o fato de o conjunto dos depositantes serem membros do corpo docente ou alunos vinculados a eles os subordina à direção do Programa DMA e, por tabela, ao presidente do Comitê Diretor; 2) o poder destes limita, consideravelmente, as atribuições do depositante e essa “blindagem” impede que avaliemos a relação do usuário com o arquivo em torno das diretrizes externas, transformando em incógnita os possíveis efeitos das situações de violação sobre os dados arquivados.
- Há mais referências ao termo “direitos autorais”<sup>113</sup> nas Diretrizes de aquisições (“Documentos relacionados ao status de direitos autorais e/ou proveniência do material”) e outra aos termos “propriedade intelectual” e “direitos éticos”<sup>114</sup> nas Políticas de aquisições (“Isso inclui os direitos de propriedade intelectual do

---

<sup>113</sup> Cf. Acquisitions guidelines. Disponível em: <<https://dma.uchicago.edu/about/acquisitions-guidelines>>. Acesso em: 20 fev. 2022.

<sup>114</sup> Cf. Acquisitions policy. Disponível em: <<https://dma.uchicago.edu/about/acquisitions-policy>>. Acesso em: 20 fev. 2022.

depositante, bem como os direitos éticos dos falantes e da comunidade de fala”), nada mais. O que se nota, portanto, nos dois documentos são referências superficiais a questões críticas de acesso na relação entre usuário e arquivo. A falta de fontes acadêmicas (livros, artigos, etc.) sobre os temas em tela nos obriga a inferir e creditar a exígua transparência das diretrizes internas e externas do acervo a dois fatos: 1) o Programa DMA limita a figura do depositante a pessoas ligadas a seu próprio meio acadêmico; 2) e a restrição ao acesso dos outros usuários somente às coleções públicas e gratuitas, com recursos limitados. O que caracteriza clara prioridade às demandas internas de pesquisa do espectro acadêmico. Porém, as características de restrição do *Digital Media Archive* contrariam duramente as bases teóricas e metodológicas da documentação linguística moderna, principalmente no que se refere à proposição de acessos completamente abertos e públicos que permitam ao usuário, pelo menos, o contato com informações básicas do corpus documental (HIMMELMANN 2007, 2006).

Após a verificação do estado da arte do Programa DMA, é possível dizer que se trata de iniciativa voltada ao desenvolvimento de projetos de documentação linguística em todos os continentes. Nesse caso, seus projetos de pesquisa são invariavelmente apoiados por grandes agências de financiamento. Como vimos, há um grande arquivo linguístico homônimo, o DMA (aberto, público e de uso limitado; carente de transparência sobre as regras de acesso e diretrizes externas, e direcionado às pesquisas do corpo docente) que armazena a maioria do acervo digital, sem que possamos considerar, entretanto, que seja o único, por haver setores na UC que armazenam o resultado das pesquisas linguísticas em repositórios próprios. Pode-se afirmar que a produção acadêmica do Programa DMA, em termos de número de projetos, pesquisas, dissertações e teses relacionadas à documentação linguística no mundo, são consideráveis. Porém, no caso do Brasil, em particular, estão em fase embrionária (ver Anexo I), tendo em vista o quantitativo dos materiais na área lançados mais recentemente. Neste caso, as iniciativas dos cursos de pós-graduação podem contribuir positivamente.

### 2.5.3 – AILLA

#### 2.5.3.1 – Antecedentes

O surgimento do *The Archive of the Indigenous Languages of Latin America* (AILLA)<sup>115</sup> em 2000, na Universidade do Texas (UT) em Austin, é resultado de iniciativa de três acadêmicos: Joel Sherzer (*Department of Anthropology*), Anthony Woodbury (*Department of Linguistics*) e Mark McFarland (*Director of the Digital Library Services Division of the University of Texas General Library System*). A criação do Programa AILLA<sup>116</sup>, responsável pelo arquivo de dados linguísticos e antropológicos das línguas indígenas da América do Sul e do Caribe, partiu da necessidade desses pesquisadores em preservar o material de suas coleções particulares (projetos de documentação, gravações, textos de línguas indígenas, anotações, etc.) de documentação das línguas indígenas. Desde então, o acesso a esse conteúdo digitalizado é realizado através do portal do Programa AILLA na Internet<sup>117</sup>.

Ao contrário do que ocorre com o ALD DMA, o arquivo AILLA não possui um correspondente físico do acervo, trata-se de um repositório totalmente digitalizado, onde as 140 coleções armazenadas são compostas por mais de 10.000 itens pesquisáveis, envolvendo 282 línguas de 22 países da América Latina e Caribe. Conforme Kung & Sherzer (2013, p. 379-383), em números precisos referentes a levantamento de 2013, essas coleções abrigam gravações de áudio (16.370), gravações de vídeo (2.155), textos digitais (4.604) e imagens (4.289) que estão em contínuo processo de crescimento.

No caso do Brasil, o resultado da busca no arquivo AILLA às coleções apresenta números significativos, são 1240 *Search results* apresentados em 25 telas de navegação (ver Anexo M). Apesar de alguns resultados da consulta não terem relação direta com o propósito de se obter corpus digital apropriado para o desenvolvimento de projetos de documentação linguística, pode-se afirmar que a maior parte do conteúdo deve atender às expectativas do pesquisador, no caso do

<sup>115</sup> Cf. Department of Anthropology. Disponível em: <<https://liberalarts.utexas.edu/anthropology/subdisciplines/Linguistic-Anthropology/ailla.php>>. Acesso em: 31 jan. 2022.

<sup>116</sup> O projeto do arquivo AILLA materializou-se por meio de bolsa fornecida pela UT e outras instituições.

<sup>117</sup> Cf. The Archive of the Indigenous Languages of Latin America (AILLA). Disponível em: <<https://ailla.utexas.org/>>. Acesso em: 31 jan. 2022.

material depositado referente às línguas indígenas brasileiras (ver Anexo N).

A maioria do acervo formou-se a partir de depósitos feitos, principalmente, por pesquisadores oriundos da América do Norte, Central, do Sul e de outros continentes, concomitantemente à criação do AILLA como projeto-piloto em 1999, e das permanentes atualizações do acervo. Sendo que a maioria dos dados arquivados é composta por gravações brutas, portanto, não analisadas de áudio e/ou vídeo. Contudo, há também transcrições, traduções e análises morfológicas interlinearizadas, de modo que ambos os formatos dos materiais disponibilizados atendem aos parâmetros formais de um corpus digital que permite o desenvolvimento de projetos de documentação linguística (ver Anexo O).

O Programa AILLA<sup>118</sup> vem identificando a existência de um espectro de línguas pouco estudadas que envolve as ameaçadas e outras com grande número de falantes, aumentando a relevância de seus projetos na América Latina. O quadro de alunos, além de ter a possibilidade de participar das pesquisas consideradas mais importantes, tem a opção de fazer essas mesmas atividades, conforme seus interesses, em qualquer parte do mundo. Além disso, há projetos recentes que se caracterizam pelo estímulo ao trabalho em equipe voltados para a DL, entre os quais destacam-se os quatro, em suma, citados abaixo:

- Projeto de documentação das línguas Maia (equipe formada por 2 docentes e 10 alunos de pós-graduação).
- Projeto de documentação da língua Chatino (equipe formada por 1 docente, 2 estudantes de pós-graduação e 4 quatro estudantes de graduação).
- Projeto dinâmica da mudança da língua de Caçadores-Coletores (equipe formada por 1 docente, 3 alunos de pós-graduação e 2 alunos de graduação).
- Projeto de documentação da língua Iquito (equipe formada por 10 alunos de pós-graduação e 4 alunos de graduação peruanos).

---

<sup>118</sup> Cf. Documentary and Descriptive Linguistics - Research Projects. Disponível em: <<https://liberalarts.utexas.edu/linguistics/graduate/concentrations/documentary/Research.php>>. Acesso em: 31 jan. 2022.

Os projetos citados ocorrem concomitantemente com os projetos em equipe. Há também os individuais realizados no campo, feitos por alunos de pós-graduação falantes nativos das línguas pesquisadas, em várias partes do mundo, com destaque àqueles desenvolvidos no Brasil com as línguas Paresi (Arawak) e Jabuti (Macro-Jê)<sup>119</sup>.

#### 2.5.3.2 – Recursos

A fim de cumprir seu objetivo principal que é preservar os recursos linguísticos e culturais das línguas indígenas latino-americanas, na maior parte em risco de extinção (KUNG & SHERZER, 2013), atualmente o AILLA se mantém por meio da junção dos esforços administrativos do LLILAS BENSON (união de instituições como *Teresa Lozano Long Institute of Latin American Studies* e *Nettie Lee Benson Latin American Collection*), Department of Linguistics University of Texas e *Digital Libraries Services Division of the University Libraries*, além dos recursos financeiros oriundos do fornecimento de bolsas de pesquisa pelo NEH e o NSF.

#### 2.5.3.3 – Arquivo linguístico digital

Em termos de método de documentação, no que tange às determinantes das melhores práticas do pesquisador em torno de sua conduta no campo de pesquisa, na relação que o envolve com auxiliares e falantes que correspondem ao conjunto do comportamento técnico e ético, não há nenhuma referência direta no portal do Programa AILLA, na Internet, nos textos consultados, no Departamento de Linguística da UT, em livros ou artigos de terceiros. Certamente, tais “negligências” têm relação direta com o fato de o acervo ser uma iniciativa voltada, exclusivamente, à recepção de material de linguistas, antropólogos, etnólogos, entre outros, para depósito no arquivo com o propósito de subsidiar dados brutos aos pesquisadores interessados na documentação das línguas ameaçadas na América Latina.

Nesse caso, pode-se inferir que as proposições dos manuais de melhores práticas para a documentação linguística no campo, como vimos, geralmente

---

<sup>119</sup>Cf. Projetos de pesquisa AILLA. Disponível em: <<https://liberalarts.utexas.edu/linguistics/graduate/concentrations/documentary/Research.php>>. Acesso em: 01 fev. 2022.

apregoadas pelas instituições acadêmicas ou pelas agências de financiamento, são implicitamente praticadas pelos depositantes no ALD do Programa AILLA. Assim, ALD o AILLA tem a função principal, como eles mesmos afirmam, de guardar uma riqueza de materiais para documentar as línguas (gramáticas, dicionários, etnografias e notas de campo)<sup>120</sup>, conforme o compromisso de servir aos povos indígenas latino-americanos e à comunidade acadêmica que atua para documentá-las. Esse objetivo do AILLA é corroborado por suas preocupações práticas em reforçar parâmetros técnicos para a continuidade de suas atividades, como sua adequação às exigências impostas para o depósito de corpora digital, com os parâmetros determinados pelos principais arquivos de documentação linguística. O resultado dessas iniciativas se reflete no crescimento de sua importância para as comunidades de fala e acadêmica.

Conforme (KUNG & SHERZER, 2013), o método de arquivamento do AILLA tem relação direta com seus princípios de criação, qual seja, a preservação dos recursos linguísticos e culturais singulares das línguas indígenas latino-americanas. Nesse sentido, assume grande relevância o depósito dos materiais documentais feitos por pesquisadores e comunidades indígenas, entre os quais destacam-se as gravações de vídeo, áudio e demais dados primários como elementos fundamentais da prática metodológica.

O repositório AILLA formado e responsável por uma diversidade de dados primários, por vezes, precisa analisar as mídias fornecidas. Todas que são analógicas e/ou estão fora do padrão aceito para o arquivamento são encaminhadas para os laboratórios próprios ou serviços de digitalização do Programa AILLA. Esse vasto conjunto de registros contém grande diversidade de gêneros discursivos (conversas, tipos variados de narrativas, músicas, discursos políticos, mitos tradicionais, cerimoniais de cura, etc.) e reúnem metadados no formato de transcrições e/ou traduções em mídias digitalizadas, ou arquivos xml. Entre os materiais textuais estão dicionários, gramáticas, esboços etnográficos, notas de campo, artigos, folhetos, apresentações e fotografias.

Contudo, o complexo armazenamento desse vasto acervo só expressa sua relevância acadêmica, linguística ou comunitária quando esses dados se tornam acessíveis ao maior número possível de interessados por intermédio da Internet. Assim, o papel dos arquivistas é fundamental para o sucesso dessa operação que

---

<sup>120</sup> Cf. La colección AILLA. Disponível em: <<https://ailla.utexas.org/es>>. Acesso em: 01 fev. 2022.

envolve, explicando de modo simplificado, a organização das etapas, a definição dos protocolos de formato de arquivos, os níveis de acesso e a implementação dos depósitos das pesquisas no arquivo AILLA. No ambiente virtual estão disponíveis, detalhadamente, os métodos de arquivamento dos materiais e as exigências gerais para sua inclusão no acervo digital<sup>121</sup>. Abaixo, estão resumidas algumas etapas desses procedimentos:

- Preencher o pacote de formulários para o depositante que incluem questões sobre os formatos de mídia, restrições de acesso, metadados, métodos de entrega e editor de metadados.
- Entregar o comprovante de consentimento informado.
- Configurar os arquivos de mídia em três formatos específicos: trabalho, apresentação e arquivo.
- Converter as mídias digitais em formatos aceitáveis pelo arquivo, conforme o padrão do Programa AILLA.

#### 2.5.3.3.1 – Código de conduta

A coleção dos arquivos multimídia AILLA está instalada em servidores administrados pelo *Digital Service of the University of Texas Libraries AILLA* (KUNG & SHERZER, 2013, p. 380). Assim como os demais arquivos digitais analisados, anteriormente, o acesso ao ALD é aberto, público e gratuito. Os interessados em consultar seu catálogo e informações congêneres podem fazê-lo sem a exigência de registro. Porém, no caso da necessidade de maiores privilégios no uso do arquivo, como download, é exigida a criação de um perfil de usuário. Ao concordar com as “Condições de uso dos Recursos Arquivados”<sup>122</sup>, o interessado se compromete a não usar comercialmente os arquivos e a respeitar a cultura dos povos e as línguas cuja singularidade está registrada nas mídias gravadas. A confirmação do cadastro permite o enquadramento do usuário em quadro níveis diferentes de acesso (ver JOHNSON, 2003 para mais informações sobre o sistema de acesso classificado):

---

<sup>121</sup> Cf. Como usar y archivar datos en el Archivo de los Idiomas Indígenas de Latinoamérica. Disponível em: <[https://ailla.utexas.org/sites/default/files/documents/como\\_usar\\_AILLA\\_2.pdf](https://ailla.utexas.org/sites/default/files/documents/como_usar_AILLA_2.pdf)>. Acesso em: 01 fev. 2022.

<sup>122</sup> Cf. Condiciones del uso de AILLA. Disponível em: <<https://ailla.utexas.org/es/site/condiciones>>. Acesso em: 20 fev. 2022.

- Acesso liberado a todos os usuários a qualquer arquivo de mídia que tenha sido depositado no nível de acesso público (nível 1).
- Acesso restrito e protegido por uma palavra-chave (nível 2).
- Acesso restrito com palavra-chave e data específica para o desbloqueio (nível 3).
- Acesso restrito por palavra-chave e controlado pelo depositante. Por e-mail, o usuário pode solicitar o acesso ao arquivo (nível 4).

#### 2.5.3.4 – Observações críticas

##### 1) Sobre o PDL AILLA

Conforme anunciado no início desta seção, as análises críticas dos programas de documentação investigados nesta dissertação, têm relação com seus esforços de responder à crise global de ameaça à diversidade linguística, denunciada nos anos 1990. A análise dessas medidas é feita a partir de três princípios estruturantes de um PDL definidos por Himmelmann (2007, p. 46–47), e consagrados por seus congêneres, como: 1) o enfrentamento à crise da diversidade linguística; 2) o aproveitamento dos recursos financeiros destinados à pesquisa documental; e 3) a verificabilidade dos dados primários por pares. Abaixo, avaliamos de que modo o Programa AILLA implementou esses princípios:

- O problema da crise da diversidade linguística foi respondido com: a) um arquivo linguístico totalmente digital para as línguas indígenas latino-americanas; b) digitalização de mídias analógicas em laboratório próprio; c) o arquivo armazena dados antropológicos e linguísticos da América Latina e Caribe.
- O aproveitamento de recursos financeiros ocorre: a) o arquivo possui centenas de coleções e milhares de itens; b) são quase três centenas de línguas arquivadas pertencentes a vinte e dois países; c) o arquivo armazena mais de mil coleções pertencentes ao Brasil.

- A verificabilidade dos dados primários é possível após: a) a preparação da análise, tradução, transcrição e anotação dos metadados; b) a transferência destes para o arquivo linguístico; c) a publicação dos recursos (corpus) linguísticos no ALD. A discussão detalhada desse tema é efetuada no tópico seguinte.

O Programa AILLA surge conforme a necessidade de seus criadores em arquivar dados primários de pesquisas linguísticas particulares. O uso do ALD criado depois se estende para os linguistas que tinham a mesma carência, mas não sabiam como resolvê-la. Por essa e outras constatações (número de projetos, quantitativo de pesquisas em andamento, volume de recursos, etc.), conclui que, em termos organizacionais e técnicos, é o programa mais transparente (em termos de compartilhamento de informações administrativas), e com estrutura de funcionamento mais adequada para linguistas que consideram criar um programa de documentação ou arquivo linguístico de médio porte.

Por fim, pelas discussões feitas no início desta seção, complementadas por essas observações, respondemos ao problema teórico desta dissertação, avaliando que o Programa AILLA cumpre, plenamente, os princípios estruturantes de um PDL ao combater a crise global da diversidade linguística, em geral, e, particularmente, a ameaça às línguas indígenas.

## 2) Sobre o ALD AILLA

O AILLA tem como pré-condições para o acesso a concordância do interessado em relação a dois documentos: as “Condições de uso AILLA” (o uso dos recursos do arquivo) e a “Licença AILLA”<sup>123</sup> (cessão de licença ao depositante) que precisam ser aceitos. Sem entrar no mérito de avaliar a totalidade dos compromissos assumidos pelo usuário ao aceitar as exigências de acesso e uso do arquivo, é preciso ressaltar que até aqui, dos três arquivos linguísticos investigados (DOBES, DMA e AILLA) sobre as regras éticas (propriedade intelectual, privacidade dos indivíduos, uso não-comercial do corpus documental e transferência de dados para arquivos externos) destacadas para analisar os pontos críticos na relação dos usuários com o arquivo

---

<sup>123</sup> Cf. Licencia de AILLA. Disponível em: <<https://ailla.utexas.org/es/site/licencia>>. Acesso em: 20 fev. 2022.

que afetam alguns direitos da comunidade, o “CdC AILLA” é o mais transparente em relação aos direitos e deveres das partes interessadas. Com isso, é possível definir como satisfatório o modo pelo qual a relação usuário-arquivo é encaminhada no âmbito daquilo que é imposto pelo código de conduta, apesar de persistirem algumas imperfeições. Assim, podemos concluir que:

- O padrão de acesso ao arquivo é aberto, público e gratuito, porém seu uso em condições privilegiadas é restrito por questões de direitos autorais. As limitações podem ser diminuídas ou incrementadas conforme a dosagem do nível de controle de acesso empregada pelo arquivista, depositante ou comunidade de fala para manter a segurança e integridade do corpus documental.
- O direito do depositante, em termos de manuseio dos materiais digitais, referendado ao assinar a “Licença do Depositante AILLA”<sup>124</sup>, ao contrário do que ocorre no DOBES, se resume a possibilidade depositar e gerenciar seu próprio corpus, pois os demais procedimentos estão a cargo dos arquivistas e técnicos do repositório. Principalmente a tarefa de administrar os pontos críticos que envolvem a segurança dos materiais digitais, conforme diretrizes internas e externas, com destaque para as seguintes autorizações: publicação dos materiais, distribuição de cópias eletrônicas, transferência de cópias a outros arquivos, permissão de uso do material para pesquisa e educação, além da liberação ou não, conforme a autorização da comunidade de fala, para o uso não comercial dos dados primários.
- A limitação do papel do depositante é acompanhada da adequada posição do arquivista a frente do gerenciamento do arquivo. Aquele que tem, entre outras funções, a responsabilidade de controlar os mecanismos de acesso, possui maior protagonismo na administração das regras éticas, algo que eleva os níveis de garantia dos direitos dos usuários e evita a possibilidade de conflitos éticos-normativos que possam comprometer um dos preceitos básicos da documentação linguística: a segurança no armazenamento do corpus digital no longo prazo.

---

<sup>124</sup> Cf. Licencia de AILLA. Disponível em: <<https://ailla.utexas.org/es/site/licencia>>. Acesso em 20 fev. 2022.

- Um ponto negativo a destacar, ao afetar a confiabilidade na relação usuário-arquivo, foi extraído da “Licença do Depositante AILLA”, em trecho no qual se afirma que no caso de “dano, perda direta, indireta ou consequente, causada pelo dano, perda ou mal uso dos materiais arquivados” o depositante além de se comprometer a assumir a responsabilidade completa sobre uma possível violação dos dados é obrigado a renunciar a qualquer acusação contra o AILLA, caracterizando certo desequilíbrio nas responsabilidades entre a parte responsável, legalmente, pela segurança e manutenção dos materiais armazenados e a outra que deposita simplesmente dados relativos à própria pesquisa.

As restrições impostas permitem aos arquivistas elevado grau de controle sobre os materiais, de modo que é possível estabelecer diferentes níveis de acesso para os arquivos individuais nas coleções ou somente um nível de acesso para toda a coleção. O que torna factível, por exemplo, liberar ao público as gravações de vídeo, bem como ocultar os nomes dos participantes da pesquisa caso desejem. Apesar desse conjunto de estratégias voltadas à segurança e integridade do acervo, o acesso dos falantes (comunidades indígenas) ao arquivo é garantido, pois como afirmam Kung & Sherzer (2013, p. 383): “[...] O AILLA é especialmente dedicado a disponibilizar a coleção para membros de comunidades indígenas da América Latina [...]”.

Por fim, é possível concluir que desde sua criação, os resultados alcançados pelo AILLA têm contribuído para além de sua proposta original, qual seja, garantir a preservação das línguas indígenas latino-americanas, na maioria ameaçadas de extinção. O desafio é grande para o arquivo que acumulou em sua base de dados mais de 140 coleções com os mais variados tipos de mídias que representam milhares de itens pesquisáveis e correspondem, simbolicamente, a uma parte da cultura produzida pela humanidade. O Brasil, a partir de uma busca fácil no repositório de línguas indígenas latino-americano, tem lugar de destaque com mais de 1000 resultados possíveis na pesquisa.

Não se pode negligenciar que o arquivo AILLA, em mais de 20 anos de existência, tornou-se uma poderosa ferramenta de estudos para linguistas e outros pesquisadores vinculados ou não ao Programa AILLA. O público interessado em geral, mas principalmente as comunidades indígenas latino-americanas, têm, sem dúvida, um instrumento de pesquisa cujo acesso é verdadeiramente, aberto, público e gratuito,

com armazenamento seguro no longo prazo - como defende veementemente HIMMELMANN (2007, 2006, 1998) e seus pares em termos teórico-metodológicos para o ALD. Essas características são extremamente importantes para aqueles que buscam meios para preservar e/ou restaurar um de seus maiores patrimônios: a língua dos ancestrais.

### 3 – ANÁLISE DE RESULTADOS

#### 3.1 – Uma abordagem crítica

A análise de resultados daquilo que vimos ao longo do processo de desenvolvimento desta dissertação envolve parte significativa das discussões que fizeram parte da composição do cenário mais atual que explica nosso objeto: “o impacto dos programas DOBES, DMA e AILLA sobre a questão das línguas indígenas ameaçadas”. Para isso, foi necessário o resgate de referenciais que abordam 4 temas específicos, divididos em 2 capítulos.

No capítulo 1 analisamos o processo de transformação da documentação linguística associado ao surgimento de novas teorias, metodologias e tecnologias que se refletiram em seu crescimento, consolidação e aplicação nas atividades de campo como subárea da Linguística em projetos de pesquisa, na maior parte, vinculados à PDL. Nesse sentido, deve-se salientar que a década de 1990 caracteriza-se por uma série de fenômenos políticos, sociais e econômicos cujos resultados não podem ser dissociados das transformações que a Linguística globalmente passou a sofrer em termos de investimentos.

Nossos estudos tornaram-se mais interessantes quando descobrimos e inferimos que fatos históricos impactantes como o colapso da União Soviética, o fim da Guerra Fria, o início da globalização (maior circulação de capitais e de pessoas), enfim, que a vitória do capitalismo global ensejou a popularização do computador pessoal, da Internet e a prosperidade econômica da Europa e dos Estados Unidos (HOBSBAWN, 2004). Tais aspectos das mudanças políticas e econômicas globais, especialmente a globalização e a incorporação forçada de mais comunidades humanas nesse sistema produtivo descentralizado, são em si fatores que contribuíram, de diversas formas, para o acelerado processo de redução da diversidade linguística.

Com esses dados, restaram menores dúvidas de que as preocupações de Krauss (1992) e demais linguísticas, naquele período, com as línguas minoritárias ameaçadas foram acompanhadas por um contexto que favoreceu um grande aporte de recursos em projetos voltados a socorrer as sociedades fragilizadas pelas ameaças a sua língua e cultura. Em linha com essas reflexões, podemos deduzir qual era a origem dos recursos que as agências de captação (governamentais ou privadas)

direcionavam para os diversos programas de documentação que surgiam, assim como as bolsas de estudo destinadas aos mestrandos e doutorandos envolvidos em pesquisas espalhadas pelo mundo. Será que hoje em dia ainda persiste essa fartura de recursos?

Todos esses investimentos não alcançariam os resultados almejados se os equipamentos e tecnologias computacionais não tivessem os custos reduzidos devido ao grande fluxo de capitais circulante provocado pelo fim das barreiras econômicas oriundas da Cortina de Ferro e do Muro de Berlim (FIORI, 2004). Enfim, seria omissão de nossa parte ignorar essas filigranas históricas e simplesmente reproduzir as motivações contidas em nossos referenciais para o *boom* dos programas de documentação linguística e sua prática, a linguística documental, ignorando que essas grandes novidades, para ocorrerem, precisaram antes da remoção de obstáculos políticos e econômicos do passado.

No capítulo 2 descobrimos que a conjunção de fatores políticos, econômicos e sociais, exposta no capítulo 1, tem relação direta com as ações existentes, atualmente, na luta contra o desaparecimento das línguas ameaçadas. No bojo desses eventos, devemos considerar a sincera preocupação da maior parte da comunidade acadêmica com as populações indígenas e com seu arcabouço linguístico e cultural, sentimento acompanhado de atitudes práticas, como o emprego da documentação linguística em atividades de pesquisa para tentar proteger a diversidade linguística.

Nesse sentido, esta investigação atesta que o século XXI apresenta uma nova realidade para as atividades relacionadas à DL. Não há mais o mesmo volume de recursos financeiros, alguns programas de documentação, como o DOBES, não captam mais recursos para novos projetos desde 2011; os interessados precisam se submeter a processos públicos de concorrência para financiamento de pesquisa e bolsas de estudo, como os divulgados pelo NSF e o NEH, por exemplo, para conseguirem subsídios para seus projetos individuais ou coletivos.

Até os PDL de maior porte que possuem fundações capazes de prover recursos próprios, atualmente em seus portais na Internet, aceitam doações, caso do DOBES. Os demais programas que também possuem ALD vinculados, como o DMA e o AILLA, captam recursos, principalmente de agências governamentais e das universidades às quais estão subordinadas. Enfim, o cenário atual é outro, mais maduro acerca das estratégias a serem adotadas em torno do problema apontado na década de 1990.

Os esforços dos PDL e dos pesquisadores talvez sejam a última fronteira de

segurança dos materiais linguísticos, analógicos ou digitais coletados no tempo presente ou em período pretérito, pois se tornaram a garantia de que outras gerações de estudiosos, povos originários e a sociedade em geral tenham assegurado um acervo apropriado para o acesso, conforme os diferentes interesses, sejam acadêmicos, de preservação ou de restauração das línguas indígenas ameaçadas ou já extintas.

Frente a necessidade de aprofundar a análise de resultados sobre nosso objeto de pesquisa e responder ao questionamento do problema deste estudo, decidimos definir os tópicos que destacam os fatores inter-relacionados que contribuíram para tornar os PDL, vinculados a seus respectivos arquivos digitais (DOBES, DMA e AILLA), agentes primordiais contrários a ameaça às línguas indígenas. Assim, dividimos as observações que julgamos essenciais sobre essa reflexão, em torno das três abordagens seguintes:

#### 1) O processo geral de constituição dos PDL

Abordar a importância conquistada pelos programas de documentação linguística ao longo das duas últimas décadas requer a ênfase de aspectos da realidade atual com o propósito de vislumbrarmos possíveis cenários futuros. Como descobrimos no capítulo 1, de modo geral, o surgimento dos PDL é o resultado de sucessivas transformações no universo social, linguístico e tecnológico transcorridas na transição dos séculos XX e XXI, entre as quais destacam-se:

- A troca das gravações analógicas feitas no campo por mídias digitais.
- O deslocamento do armazenamento das mídias de locais físicos para o ambiente virtual.
- A revisão teórica e metodológica da documentação linguística, habilitando-a aos novos recursos tecnológicos e, por conseguinte, elevando-a, definitivamente, a subárea da Linguística.
- A confluência de três fatores interligados: o agravamento da ameaça às línguas minoritárias; a mobilização de instituições governamentais, privadas e setores acadêmicos; e os fortes investimentos em projetos de documentação linguística e arquivamento.

- O depósito das coleções desenvolvidas na pesquisa de campo nos primeiros arquivos linguísticos digitais, evidenciando a aproximação entre as ciências humanas e as ciências exatas através da TIC.

Nesse cenário, deve-se destacar os esforços dos PDL e acadêmicos na elaboração de teorias e metodologias que tornaram a DL instrumento crucial em todo o processo de documentação, desde a gravação inicial até a disponibilização da coleção na Internet (AUSTIN, 2010; HIMMELMANN, 2006). E como qualquer ciência que avança por critérios dialéticos, visando seu aprimoramento, surgiram variadas críticas em relação à teoria, métodos e à cientificidade de seus procedimentos (MOORE & GALUCIO, 2016; GRENOBLE & FURBEE, 2010).

Em meio à complexidade de todas as situações elencadas no processo de constituição e consolidação geral dos PDL, apresentamos o último quesito sobre as breves análises dos programas de documentação linguística investigados.

## 2) Os fundamentos dos PDL DOBES, DMA e AILLA

Nesta etapa faremos breve abordagem dos pontos centrais destacados da dissertação referentes aos principais elementos estruturantes dos programas de documentação linguística analisados.

### DOBES:

- O Programa DOBES apresenta a mais robusta estrutura organizacional. Essa característica que se reflete em montante de recursos investidos; número de projetos desenvolvidos; replicação de seu modelo de documentação e arquivos digitais; produção de teorias e metodologias próprias; criação de banco de dados linguístico; e desenvolvimento de softwares para a análise de dados, entre outras iniciativas.
- A qualidade e padrão de funcionamento levou outros acervos a replicar elementos teóricos, metodológicos e tecnológicos do PDL DOBES na América Latina, como o PRODOCLIN e o Museu Goeldi no Brasil; o DILA na Argentina; e o IIAP no Peru. Não podemos ignorar que essa expansão tem relação com o contexto histórico citado anteriormente, e com a necessidade de acelerar o

processo de documentação nos locais onde havia maior diversidade linguística (México, Guatemala e Brasil).

- O fato de ser um dos PDL com maior volume de recursos investidos reflete-se no porte de sua estrutura tecnológica de arquivamento, capaz de armazenar materiais linguísticos digitais coletados de várias origens na matriz holandesa ou localmente, nos países onde os projetos foram feitos.
- Na relação com o usuário, o PDL DOBES é omissivo em pelo menos três situações: 1) não prioriza o acesso do público a seu conteúdo mais avançado, restringindo-o a pesquisadores, consultores e comunidades de falantes, algo que denota uma conduta corporativista em termos de compartilhamento dos materiais arquivados. O pesquisador (depositante) que detém privilégios nos controles de acesso, consegue autorizar a entrada no arquivo e uso da coleção por usuário exógeno ao grupo responsável pela coleção, ação que rompe com os princípios de direitos de propriedade intelectual e privacidade, além de submeter o ALD a riscos de integridade de dados; 2) as ações deletérias do depositante em relação ao arquivo refletem a omissão do PDL DOBES sobre as competências do arquivista e do técnico com reflexo, mais uma vez, sobre a segurança das coleções digitais disponíveis no ambiente web. 3) Por fim, não localizei, nos referenciais sobre o PDL, apesar de seu porte organizacional, qualquer menção às medidas de cibersegurança<sup>125</sup> sobre seu arquivo linguístico digital.

---

<sup>125</sup> Conforme Steinberg (2006): “a cibersegurança visa fornecer a confidencialidade, integridade e disponibilidade dos dados. A confidencialidade é a garantia de que as informações não serão divulgadas ou disponibilizadas a entidades não autorizadas; a integridade é a garantia de que os dados são precisos e completos, ou seja, não sofreram nenhum tipo de violação; a disponibilidade é a garantia de que as informações, os sistemas usados para o armazenamento e processamento, os mecanismos de comunicação usados para acesso e retransmissão e todos os controles de segurança associados funcionam corretamente.”

## DMA:

- As descobertas feitas em torno do Programa DMA, vinculado à UC, nos permitem afirmar que se trata de PDL de médio porte se comparado, por exemplo, ao PDL DOBES. Conforme os levantamentos, possui mais de 2000 coleções cobrindo quase 300 línguas e projetos desenvolvidos e em andamento em todos os continentes. Sua estrutura é sustentada mediante doações e por agências de financiamento governamentais e privadas.
- Entre as novidades a destacar sobre o Departamento de Linguística da UC, estão sua relação com os projetos de documentação linguística direcionados às línguas ameaçadas, a existência, desde 1950, do espaço físico destinado ao armazenamento de materiais analógicos voltados a estudos multidisciplinares e sua histórica relação com o Departamento de Estado norte-americano durante a Segunda Guerra Mundial e a Guerra Fria.
- O arquivo linguístico digital do PDL DMA, se comparado aos ALD DOBES e AILLA, é aquele com maior restrição de informações públicas sobre o seu funcionamento. De modo que situações como a relação do usuário com o arquivo são desconhecidas. Portanto, ficaram dúvidas se alguns dos principais direitos dos usuários (propriedade intelectual, privacidade, uso não comercial, etc.) estão sendo respeitados, tendo em vista que os depósitos e o acesso ao arquivo são controlados e fiscalizados somente por seus diretores, sem que o público externo tenha conhecimento de seus mecanismos.
- A lista de omissões do PDL DMA se destaca em relação aos outros pontos discutidos de modo que, pelo menos, quatro observações precisam ser feitas: 1) a primeira se refere a inexistência de publicações que abordem os princípios teóricos e metodológicos que norteiam o desenvolvimento dos projetos de documentação linguística, algo que interdita qualquer iniciativa que pretenda destacar os princípios que regem as atividades de coleta, análise linguística, digitalização, depósito/arquivamento e publicações para o acesso às coleções digitais; 2) em segundo lugar, está a negligência do programa quanto às políticas de acesso ao arquivo pelo usuário, na maior parte dos casos, limitado ao nível 1, permitindo somente a visualização de metadados do corpus linguístico. De modo que, mesmo no caso de interesse, o cadastro exigido para os níveis mais avançados de uso do ALD raramente resulta em resposta dos

administradores ligados ao programa. O próprio responsável por esta dissertação foi ignorado nessa tentativa; 3) o portal do PDL DMA e da UC, na Internet, representam a fonte de todas as nossas referências e, como vimos, isso ocorreu devido à falta de publicações de outro tipo. Essa negligência também se refletiu na quase nulidade de informações sobre os projetos conduzidos na América Latina, em termos de número, dimensão e locais; 4) finalmente, assim como o Programa DOBES, deve-se ressaltar que o Programa DMA também não divulga a existência de qualquer política de cibersegurança para evitar riscos de integridade a seu arquivo linguístico<sup>126</sup>.

#### AILLA:

- As pesquisas sobre o Programa AILLA revelaram sua dependência de recursos financeiros provenientes da UT, de agências de financiamentos governamentais norte-americanas e de outros países. Suas 140 coleções são formadas por quase 300 línguas de 22 países da América Latina e Caribe, destacando-se o número de coleções referentes ao Brasil e às línguas indígenas brasileiras.
- O ALD, totalmente digital, tem em vista disponibilizar suas coleções para as comunidades indígenas da América Latina. A coleção é alimentada por doações de mídias analógicas que, após a digitalização, são incorporados ao restante do acervo.
- O código de conduta do PDL AILLA é o mais transparente em relação às regras de acesso ao arquivo, envolvendo pesquisadores, consultores e comunidade de fala. Entre as novidades apresentadas pelo ALD, destaca-se que o arquivista do programa é o responsável pelo controle de acesso ao arquivo, restringindo a atuação do depositante sobre a operação.

---

<sup>126</sup> A World Wide Web tornou-se um ambiente onde constantemente corporações que lidam com volumes gigantescos de informação têm seus aparatos de segurança constantemente testados, pois ciberataques de criminosos virtuais buscam, o tempo todo, fragilidades nos sistemas de informação instalados que permitam o roubo ou sequestro de dados valiosos (financeiros, tecnológicos, biológicos, acadêmicos, etc.), situação que transformou variados tipos de informação sensíveis em verdadeiras commodities. Disponível em: < <https://elpais.com/tecnologia/2022-04-08/los-ciberataques-se-multiplican-desde-la-invasion-de-ucrania-y-se-hacen-mas-frecuentes-diversos-y-complejos.html>>. Acesso em: 08 abr. 2022.

- Como vimos, o ALD AILLA apresenta seus métodos de acesso e arquivamento de forma transparente, mas deixa dúvidas sobre sua prática documental até mesmo na obra do seu mentor (KUNG & SHERZER, 2013). A explicação para essa negligência pode residir em dois fatores: a falta de melhor interlocução com a equipe de pesquisadores da UT ou a utilização de manuais elaborados por outras instituições com instruções sobre melhores práticas para essa atividade de campo.
  - Em relação à omissão do PDL sobre sua responsabilidade por danos ao arquivo e a segurança na Internet, abordamos dois pontos importantes: 1) a observação negativa em relação ao Código de Conduta AILLA trata do desequilíbrio de poder no caso de violação de dados linguísticos. O programa, nessa situação, imputa somente ao depositante a responsabilidade pelo dano, proibindo-o de qualquer acusação contra o arquivo, ou seja, isentando-se de qualquer responsabilidade; 2) em termos de política de cibersegurança, o PDL AILLA reproduz inação semelhante aos demais programas de documentação analisados, ou seja, não há nenhum tipo de proposta nesse sentido.
- 3) Síntese do processo de consolidação dos programas de documentação e a resposta ao problema da pesquisa

Após a indicação e análise dos elementos centrais das reflexões que sustentam as temáticas levantadas nos capítulos 1 e 2, apresentamos o seguinte e definitivo quadro sintético: 1) do início dos anos 1990 até a primeira década do século XXI tem-se um cenário político, econômico e social favorável a investimentos diversificados de capital, combinado com o avanço e barateamento dos recursos tecnológicos; 2) desde então, até os dias atuais, a instabilidade econômica global (explicadas por crises cíclicas do capitalismo, como a ocorrida mais recentemente, em 2008), provoca a escassez de recursos financeiros que se reflete em diversas áreas, entre elas, a acadêmica.

Em relação ao problema (teórico) desta dissertação, levantado no capítulo 1, e às respostas dadas à questão da crise da diversidade linguística, informamos que os programas (ver capítulo 2, seção 2.5) foram submetidos à verificação do cumprimento dos três princípios estruturantes de um PDL definidos por Himmelmann (2007, p. 46–47), e consagrados por seus congêneres. Estes são: 1) o enfrentamento à crise da

diversidade linguística; 2) o aproveitamento dos recursos financeiros destinados à pesquisa documental; e 3) a verificabilidade dos dados primários por pares. Os resultados da avaliação do cumprimento de cada um deles foram plenos para os PDL DOBES e AILLA; o DMA foi exceção (princípio da verificabilidade). Entretanto, esse resultado, como vimos, não invalida as outras iniciativas do PDL. Assim, podemos afirmar que o impacto dos três programas de documentação no enfrentamento da ameaça à diversidade linguística, com destaque para as línguas indígenas, é altamente positivo e produtivo.

Alinhados com as constatações acima, sublinhamos que os ALD simbolizam o maior poder dos PDL, pois seus recursos inimagináveis de armazenamento e compartilhamento de tesouros digitais pertencentes a organizações públicas ou privadas permitem a disponibilização do corpus linguístico para setores da sociedade solidários a causa ameaça à diversidade linguística. Deste modo, com este e outros objetivos de salvaguarda do patrimônio imaterial humanidade, apresentamos uma proposta básica de projeto de arquivo linguístico digital para o Setor de Linguística do Museu Nacional/UFRJ que segue os parâmetros listados nas páginas seguintes.

### 3.2 – Arquivo linguístico digital do Setor de Linguística/MN/UFRJ

Antes de apresentarmos os passos essenciais direcionados à elaboração de uma proposta básica de projeto de banco de dados de um arquivo linguístico digital para o Setor de Linguística/MN/UFRJ, destacamos algumas características, em termos administrativos e técnicos, que esse sistema deve possuir para serem evitados os problemas detectados, apresentados e criticados na análise de resultados, principalmente aqueles, repetidamente, encontrados nos três programas de documentação de modo que se desenvolva um produto adequado:

- Em termos administrativos, a criação de um Código de Conduta que proponha políticas de acesso bem elaboradas e claras que não permitam riscos às coleções e estabeleçam equilíbrio na relação usuário-arquivo quanto a responsabilidade sobre danos ao material arquivado.
- Em termos técnicos e de segurança, elaborar uma política de implantação robusta para o ALD na Internet; bem como seguir as melhores práticas sobre cibersegurança em um contexto em que os crimes virtuais atingem dimensões alarmantes.
- Em termos de local para a instalação dos aplicativos, deve-se utilizar o ambiente corporativo da Intranet para abrigar o futuro arquivo linguístico digital SLMN por questões de segurança e privacidade.
- Em termos de custo (gratuito) e atualização constante, sugerimos a utilização de software GPL<sup>127</sup> no processo de construção do banco de dados, nesse caso recomendamos o *MYSQL*<sup>128</sup> como Sistema de Gerenciamento de Banco de Dados (SGBD). E como software para a elaboração do produto (aplicativo front-

---

<sup>127</sup> “GNU is a project led by Richard Stallman of the Free Software Foundation. GNU stands for “Gnu is Not UNIX”; the g is usually pronounced, like “g’noo.” The GNU project’s original goal was to develop a freely distributed substitute for UNIX, but Linux (the leading free UNIX) evolved into a somewhat separate project. The most important GNU products are the Emacs editor and the GNU C Compiler (DOWNING, 2013, p. 215)”.

<sup>128</sup> Banco de dados de código aberto (GPL) líder mundial, usado pelos maiores provedores de nuvem e empresas da web, incluindo Facebook, Twitter, eBay, Pinterest, Tumblr, Wikipedia e YouTube. Está incorporado às soluções de mais de 3.000 fornecedores de software/OEMs independentes, incluindo Adobe, Alcatel-Lucent, Cisco, General Electric, Hewlett-Packard e Symantec. Disponível em: < <https://www.oracle.com/webfolder/college-recruiting/projects/mysql.html> >. Acesso em: 22 abr. 2022.

end<sup>129</sup>), usado na interface com o usuário e também GPL, indicamos a utilização do PHP<sup>130</sup> para a criação da *web page* (NIEDERAUER, 2017).

Para a proposta básica de projeto de arquivo linguístico para o Setor de Linguística/MN/UFRJ, nos guiamos pelos ALD estudados nesta dissertação e nos referenciais que apresentaram as melhores práticas, tecnologias e teorias destinadas ao gerenciamento e construção de banco de dados (BD). Em torno desses objetivos, dividimos essa tarefa em três partes: 1) conceitos introdutórios, 2) visões de banco de dados e 3) projeto de banco de dados.

### 1) Conceitos introdutórios

Entre os usuários interessados em ingressar no mundo dos bancos de dados é necessário o reconhecimento de diversos conceitos relativos ao assunto, porém devido às limitações daquilo que nos propomos nesta seção, nos limitaremos a apresentação daqueles conceitos fundamentais a seu desenvolvimento, entre os quais: dado, informação, conhecimento, banco de dados e sistema de gerenciamento de banco de dados.

Conforme Teorey et al. (2014), no ambiente de desenvolvimento tecnológico direcionado ao desenvolvimento de bancos de dados, conceitos como dado, informação e conhecimento assumem os seguintes significados:

- Dado é o elemento básico do arquivo com correspondência no mundo real (minimundo), formando um sistema de arquivos, como: nome, sobrenome, endereço, bairro, cep, etc.
- Informação é o resultado da análise dos dados mediante hipóteses que podem ou não ser refutadas, ou por meio de correlação de dados que apresentem hipóteses promissoras passíveis de consideração.

---

<sup>129</sup> Conforme Takai (2005, p.11): "Em arquitetura de banco de dados Cliente-Servidor, o cliente (front-end) é o responsável pela execução do aplicativo, é quem fornece a interface do usuário, ou seja, as telas que permitem o processamento da entrada e saída de dados."

<sup>130</sup> "PHP is a scripting language often used to generate web pages by computation on the server (compare CGI, definition 1; ASP). PHP originated as a set of small, efficient CGI programs to do common tasks, released by Rasmus Lerdorf in 1995. The package was called PHP Tools, short for Personal Home Page Tools. Today, PHP is said to stand for PHP Hypertext Processor (yes, the acronym includes itself; compare GNU). PHP is available free of charge for a wide range of computers." (DOWNING, 2013, p. 364)

- Conhecimento é a capacidade adquirida pelo usuário de discernir, criteriosamente, sobre os dados e informações disponíveis.

Portanto, podemos dizer que o dado é o elemento bruto que compõe o arquivo, enquanto o registro é a representação da informação, pois se conforma pela sequência de dados reunidos, como, por exemplo, as fichas cadastrais de um sistema bancário (MACHADO, 2008).

Em uma rápida consulta aos referenciais teóricos relativos à área de desenvolvimento de sistemas, e, especificamente, a bancos de dados, verificamos que existem algumas definições conceituais sobre o tema:

- “Basicamente nada mais é do que um sistema de armazenamento de dados baseado em computador; isto é, um sistema cujo objetivo global é registrar e manter informação. Esta informação pode ser qualquer uma considerada significativa à organização servida pelo sistema - em outras palavras, qualquer uma necessária ao processo de decisão da gerência daquela organização (DATE, 1989, p. 26).”
- “Um banco de dados é um objeto mais complexo; é uma coleção de dados armazenados e inter-relacionados, que atende às necessidades de vários usuários em uma ou mais organizações, ou seja, coleções inter-relacionadas de muitos tipos de tabelas [...]” (TEOREY et al., 2014, p. 2).

Podemos resumir as definições acima, afirmando que os banco de dados são sistemas informáticos responsáveis pelo armazenamento de dados relacionáveis entre si, e importantes para as organizações poderem disponibilizá-los aos usuários para a tomada de decisões ou para coleta de dados que proporcionem informações capazes de produzir o conhecimento.

Os sistemas de gerenciamento de banco de dados são softwares responsáveis pelo gerenciamento do banco de dados em diversos níveis. Conforme Teorey et. al (2014), o SGBD permite manipular, definir e construir um banco de dados para diversas aplicações. Sendo composto de uma ferramenta lógica capaz de projetar um BD, linguagens de definição e manipulação de dados e vários softwares importantes. Entre os mais utilizados estão: *Oracle Database, MySQL, Microsoft SQL Server, PostgreSQL, MongoDB*, entre outros<sup>131</sup>.

---

<sup>131</sup> Cf. “10 Sistemas de Banco de Dados mais usados atualmente.” Disponível em: <<http://www.bosontreinamentos.com.br/bancos-de-dados/10-sistemas-de-bancos-de-dados-mais-usados-atualmente/>>. Acesso em: 10 abr. 2022.

## 2) Visões de um Banco de Dados

Conforme Silberschatz et al. (2006), o sistema de banco de dados é uma coleção de dados inter-relacionados e um conjunto de programas que permitem aos usuários acessar e modificar esses dados. Tem como um dos objetivos mais importantes fornecer aos usuários uma visão abstrata dos dados, com isso, o sistema oculta certos detalhes de como os dados são armazenados e mantidos. SGBD seguros e eficientes são complexos e contam com a figura do Administrador de Banco de Dados (em inglês, *DBA*<sup>132</sup>), profissional cuja função é auxiliar o usuário, segundo Mullins (2002), garantindo a (o):

- Design de banco de dados.
- Monitoramento e ajuste de seu desempenho.
- Disponibilidade contínua.
- Segurança.
- Backup e recuperação.
- Integridade de dados.

Além disso, também é tarefa do DBA realizar abstrações sobre os dados a fim de sistematizar suas atividades profissionais em três níveis, como atesta Silberschatz et al. (2006, p. 4):

- “Nível físico: O nível mais baixo de abstração descreve como os dados são realmente armazenados. O nível físico descreve detalhadamente as complexas estruturas de dados de baixo nível.”
- “Nível lógico: O nível imediatamente mais alto de abstração descreve quais dados são armazenados no banco de dados e quais relações existem entre esses dados. O nível lógico, portanto, descreve todo o banco de dados em termos de um pequeno número de estruturas relativamente simples.”
- “Nível visões: O mais alto nível de abstração descreve apenas parte do banco de dados [...] Muitos usuários do sistema de banco de dados não precisam de todas essas informações; em vez disso, eles só precisam acessar uma parte do banco de dados. O nível de abstração de visão existe para simplificar sua interação com o sistema. O sistema pode fornecer muitas visualizações para o mesmo banco de dados.”

---

<sup>132</sup> Significado da sigla em inglês: *data base administrator*.

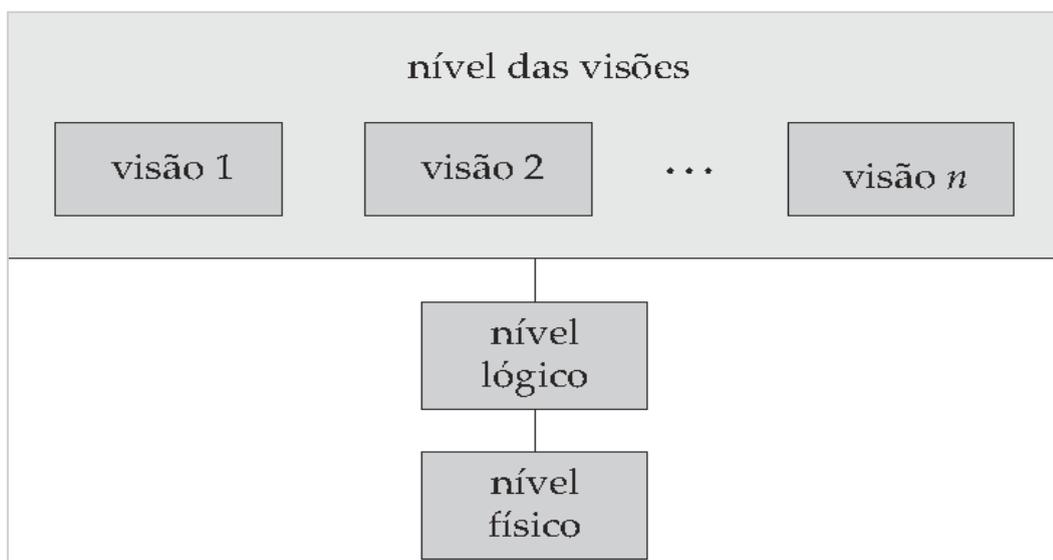


Figura VIII: Níveis de abstração de dados<sup>133</sup>

### 3) Projeto básico de Banco de Dados

Os sistemas de gerenciamento de banco de dados possuem um ciclo de vida para a sua execução. O ciclo de vida se refere às etapas sistemáticas necessárias para o projeto de um banco de dados que envolve o projeto lógico, independente do ambiente do sistema, até o projeto físico, baseado nos detalhes do SGBD escolhido para implementar o BD. Conforme Elmasri & Navathe (2005, p. 261), esse ciclo de vida se divide segundo as 8 atividades destacadas:

- “Definição de sistemas: aqui são definidos o escopo do sistema de banco de dados, seus usuários e suas aplicações. São identificadas as interfaces para as várias categorias de usuários, as restrições para tempos de resposta e as necessidades de armazenamento e processamento.”
- “Projeto de banco de dados: Ao término desta fase, o projeto lógico e físico completo do sistema de banco de dados, no SGBD escolhido, estará pronto.”
- “Implementação do banco de dados: Inclui a especificação das definições conceituais, externa e interna do banco de dados, criando o banco de dados com arquivos vazios e implementando as aplicações de software.”
- “Carregamento ou conversão do banco de dados: O banco de dados é carregado por meio da inserção direta dos dados ou pela conversão, para o formato do sistema, de arquivos existentes.”
- “Conversão das aplicações de software: Qualquer software aplicativo do sistema anterior é convertido para o novo sistema.”
- “Teste e validação: O sistema novo é testado e validado.”
- “Operação: O sistema de banco de dados e suas aplicações são colocados em operação. Normalmente, o novo sistema e o antigo operam paralelamente durante algum tempo.”

<sup>133</sup> Figura retirada e adaptada de Silberschatz et al (2006, p. 7).

- “Monitoramento e manutenção: durante a fase operacional, o sistema é constantemente mantido e monitorado. Pode haver crescimento e expansão de conteúdo de dados e aplicações de software. Ocasionalmente, podem ser necessárias modificações maiores e reorganizações.”

Nos remetemos ao segundo item da lista do ciclo de vida de um SGBD (“Projeto de banco de dados”), visto acima, para tratarmos dos conceitos de modelo conceitual, lógico e físico. Tais modelos têm a finalidade de registrar as informações coletadas do mundo real, permitindo a criação de um “minimundo”, cujas “visões” subsidiarão os conhecimentos do analista/desenvolvedor, validados conforme os objetivos e metas estabelecidos para a construção de um sistema de banco de dados consoante a necessidade do usuário:

- “O modelo conceitual é uma descrição do banco de dados de forma independente de implementação em um SGBD. O modelo conceitual registra que dados podem aparecer no banco de dados, mas não registra como estes dados estão armazenados ao nível de SGBD [...]” (HEUSER, 2004, p. 25).
- “O Modelo Lógico tem seu início a partir do Modelo Conceitual, considerando uma das três abordagens dos SGBD atualmente possíveis: relacional, hierárquica e de rede [...]” (MACHADO, 2004, p. 25).
- “O Modelo Físico irá partir do Modelo Lógico e descreve as estruturas físicas de armazenamento de dados, tais como: tamanho de campos, índices, tipo de preenchimento destes campos, nomenclaturas, etc., projetadas conforme os requisitos de processamento e uso mais econômico dos recursos computacionais. Este modelo detalha o estudo dos métodos de acesso do SGBD, para elaboração dos índices de cada informação colocada nos Modelos Conceituai e Lógico [...]” (DATE, 2004, p. 30).

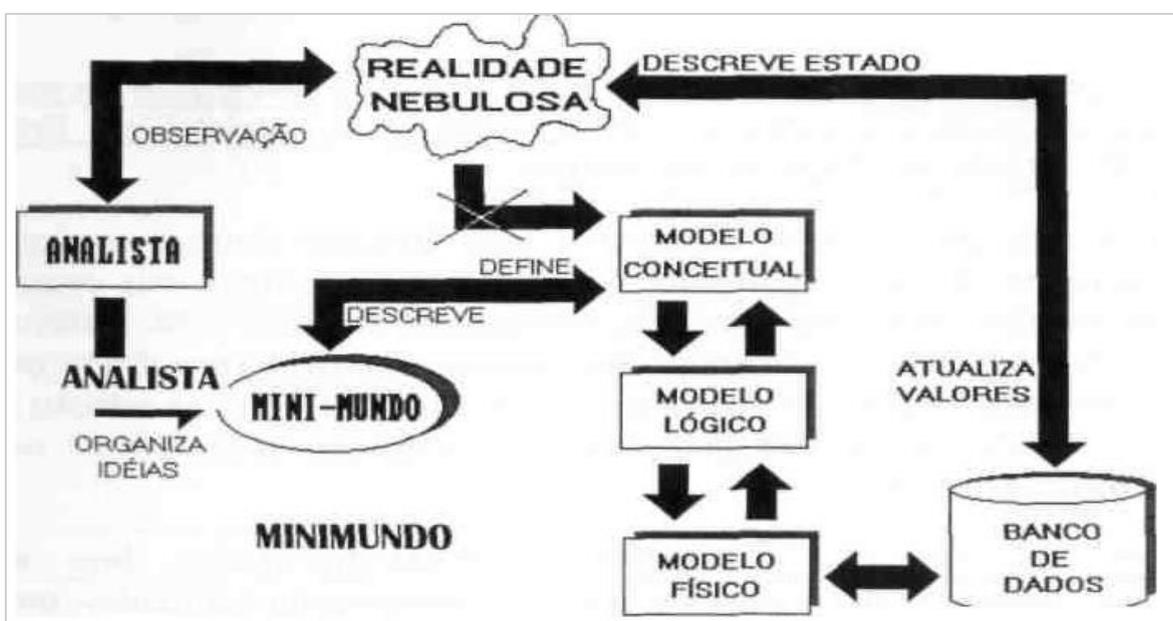


Figura IX: Abstração da coleta de requisitos e da sequência de aplicação dos modelos<sup>134</sup>

<sup>134</sup> Figura retirada e adaptada de Machado (2004, p. 22).

Todas as etapas estabelecidas e discutidas nesta proposta básica de um projeto de arquivo linguístico digital para o SLMN abrangeram os conceitos essenciais que envolvem banco de dados e projetos destinados à sua criação (ciclo de vida: modelo conceitual, lógico e físico) ao considerar as funções primordiais para o desenvolvimento e gerenciamento dessa ferramenta. Com base nesses pressupostos, apresentamos abaixo, nossa proposta básica de projeto de arquivo linguístico digital.

### 3.2.1 - Estrutura do projeto

Para complementar e aprofundar as informações correspondentes a nosso de projeto básico de banco de dados, principalmente aquelas que remetem a seu ciclo de vida (modelo conceitual, modelo lógico e modelo físico), informamos que esses elementos estruturam as etapas necessárias para a construção de uma aplicação web. A demonstração de cada um desses passos, da prática profissional, refere-se ao desenvolvimento de uma ferramenta de gerenciamento das coleções de documentação produzidas pelos pesquisadores do Setor de Linguística do Museu Nacional (SLMN) a serem armazenadas no arquivo digital (banco de dados) próprio. Assim, Alves (2017, p. 47) destaca a importância da implementação das seguintes etapas no processo de prototipação<sup>135</sup>:

- Descrever o software a ser construído
- Levantar os objetivos e requisitos do sistema
- Especificar as funcionalidades do sistema
- Definir o Diagrama de Contexto
- Preparar a lista de eventos
- Estruturar o banco de dados
- Desenvolver interface com usuário

---

<sup>135</sup> A prototipação refere-se à criação de modelo funcional de sistema de computador ou software para testagens e refinamentos. É geralmente empregada com equipamentos e programas específicos para o desenvolvimento de novos sistemas computacionais (hardware/software) e nas aplicações voltadas ao gerenciamento de informações, como os bancos de dados (SAWAYA, 1999).

### 3.2.2 – Apresentação

Nesta apresentação da modelagem do sistema, informamos que algumas políticas de restrição ao usuário ocorrem até que ele defina seu perfil. A partir do momento que esse perfil é determinado (usuário externo, depositante, arquivista, etc.), o nível atribuído a ele tem a função de balizar o uso do sistema. No caso específico do usuário externo, assumindo como critério a segurança e integridade da aplicação, será fornecido um formulário eletrônico cujo preenchimento possibilitará traçar o perfil e nível deste público (acadêmico, pesquisador, estudante, etc.) de modo que seja possível controlar suas atribuições de uso. Ou seja, o conteúdo acessível ao público externo será definido pelas informações prestadas através do formulário sem envolver material restrito. Abaixo seguem os detalhes técnicos dessa política de acesso.

Em nosso de projeto básico de desenvolvimento de um arquivo linguístico no ambiente web para o SLMN, dividimos o software nos módulos arquivista, depositante e usuário (público), onde cada um possui diferentes funcionalidades e modos de acesso, conforme as diretrizes descritas abaixo:

- O acesso ao módulo arquivista permite, através da utilização de senha de acesso com permissões de nível máximo, a ingerência sobre os demais módulos. Assim, o arquivista tem autorização para: 1) cadastrar, editar e consultar depositante, usuário e coleção; 2) depositar novo recurso na coleção.
- O acesso ao módulo depositante autoriza, pelo uso de senha de acesso com permissões de nível intermediário (com restrições), o depositante a: 1) cadastrar-se, editar e consultar os próprios dados; 2) cadastrar, editar e consultar sua coleção; 3) depositar novo recurso em sua coleção,
- O acesso ao módulo público permite ao usuário: consultas simples e downloads de algumas mídias documentais liberadas pelo arquivista ou depositante. No caso de acessos restritos, será necessário o login com uso de senha com permissões de nível básico (maiores restrições), permitindo o cadastro do usuário, a edição dos seus dados, a consulta da coleção e o download de alguns recursos restritos liberados pelo depositante/arquivista.

Encerrada a apresentação do projeto, passamos para o detalhado

levantamento dos requisitos da aplicação<sup>136</sup> que se reflete no desenvolvimento da primeira parte de sua modelagem<sup>137</sup>.

### 3.2.3 – Identificação dos objetivos

Este projeto define-se, basicamente, pelos objetivos gerais direcionados à elaboração de um arquivo linguístico. Tais objetivos são bastante semelhantes aos levantados junto aos arquivos linguísticos metodicamente analisados nesta dissertação (DOBES, DMA e AILLA), entre os quais o armazenamento de documentação linguística resultante de pesquisas de campo referentes às línguas indígenas, permitindo sua documentação.

Em termos técnicos e tecnológicos, acrescentamos às funcionalidades básicas já apresentadas no tópico anterior, a identificação detalhada dos objetivos e aspectos mais relevantes para a fundamentação de nosso projeto ao descrever suas características e recursos, conforme aquilo que é textualmente listado abaixo:

- O software deve possuir um módulo arquivista, de acesso exclusivo, a partir do qual seja possível o cadastro, edição e consulta de dados do depositante, do usuário e da coleção.
- A aplicação deve possuir um módulo depositante, de acesso permitido ao arquivista e ao depositante, de modo que a este seja permitido incluir, alterar e consultar os próprios dados e incluir, alterar e consultar os dados da coleção.
- Em resumo, enquanto o módulo arquivista, de uso restrito, tem o papel de organizar e gerenciar todo o ALD, o módulo depositante é responsável pela administração da coleção de cada depositante.

---

<sup>136</sup> Conforme Elmasri & Navathe (2005, p. 9): "... trata-se da tarefa na qual o projetista de banco de dados identifica os dados que devem ser armazenados e define as estruturas adequadas de armazenamento de dados. Essas atividades são feitas antes que o banco seja efetivamente criado e depende do contato desse profissional com os futuros usuários a fim de levantar seus requisitos e iniciar a elaboração do projeto."

<sup>137</sup> A modelagem conceitual de uma aplicação se refere a configuração do banco de dados e dos programas a ele associados, criados para efetuarem cadastros, edições, consultas e atualizações. Esse processo geralmente ocorre por meio de interface gráfica - *User-Friendly Graphical User Interfaces* (GUIs) - apresentada no padrão de formulários para a interação com o usuário final (ELMASRI & NAVATHE, 2005).

- A aplicação também deve possuir o módulo público, cujo acesso é aberto na web<sup>138</sup>, tendo como front-end a tela principal do site. Ao acessar o ALD, o usuário pode consultar e efetuar downloads de parte dos recursos<sup>139</sup> de uma coleção liberados pelo depositante ou arquivista. No caso de coleção com restrição de acesso, é necessário o cadastramento do usuário e a autorização de acesso a ser concedida pelos mantenedores.

Como vimos, os módulos representam os perfis (arquivista, depositante e público) que estão aptos a manipular o banco de dados linguístico. Dando continuidade à modelagem, vamos dispor as informações apresentadas acima, relacionando aspectos como funcionalidade, permissão e público-alvo, conforme os quadros apresentados abaixo:

Funcionalidade	Permissão	Público-alvo
Cadastro de depositante	Inclusão, edição e consulta de dado do depositante	Arquivista
Cadastro de usuário	Inclusão, edição e consulta de dados do usuário	Arquivista
Cadastro de coleção	Inclusão, edição e consulta de dados da coleção	Arquivista
Cadastro depositante	Inclusão, edição e consulta de dado do depositante	Depositante
Cadastro de coleção	Inclusão, edição e consulta de dados da coleção	Depositante

Quadro VII: Funcionalidades, permissões e público-alvo dos módulos arquivista e depositante<sup>140</sup>

Funcionalidade	Permissão	Público-alvo
Consulta de coleção	consulta à coleção e download de recursos liberados	Usuário externo
Cadastro de usuário	Inclusão, edição e consulta de dados do usuário e às coleções restritas conforme o nível de acesso	Usuário externo

Quadro VIII: Funcionalidades, permissões e público-alvo do módulo público<sup>141</sup>

Nesta etapa final de refinamento<sup>142</sup> do levantamento das necessidades do

<sup>138</sup> Apesar de fazer referência ao termo “web” (World Wide Web) ou Internet como possíveis locais onde a aplicação pode ser publicada e utilizada, ressaltamos o que afirmamos em nossa recomendação exposta na análise de resultados, qual seja, o ideal, por questões de cibersegurança, é que o software seja implantado na intranet.

<sup>139</sup> No processo de depósito na coleção específica, como, por exemplo, em arquivo linguístico como o AILLA, um recurso equivale ao armazenamento de vários arquivos documentais agrupados. Disponível em: < [https://ailla.utexas.org/sites/default/files/documents/como\\_usar\\_AILLA\\_2.pdf](https://ailla.utexas.org/sites/default/files/documents/como_usar_AILLA_2.pdf)>. Acesso em: 17 out. 2022.

<sup>140</sup> Quadro retirado e adaptado de Alves (2017, p. 49).

<sup>141</sup> Quadro retirado e adaptado de Alves (2017, p. 49).

<sup>142</sup> O refinamento é a decomposição do problema em níveis cada vez mais detalhados. Essa norma é aplicada à divisão das funções de uma aplicação (modularização) de modo que seja possível a elaboração de algoritmos e codificação em linguagem de programação (ALVES, 2017, p. 26).

sistema, utilizamos as informações disponibilizadas nos quadros acima para aprofundar o papel de cada uma das funcionalidades citadas e destacar seus modos de execução, com foco, primeiramente, nos módulos arquivista e depositante:

- Cadastro de depositante – O arquivista cadastra novos depositantes, além disso, é permitida a edição e a consulta de dados de depositantes cadastrados.
- Cadastro de usuário – O arquivista cadastra novos usuários, além disso, é permitida a edição e consulta de dados de usuários cadastrados. No cadastro é definido o nível de acesso do usuário (1 até 4).
- Cadastro de coleção – O arquivista cadastra novas coleções, além disso, é permitida a edição e consulta de dados de coleções cadastradas.
- Cadastro de depositante – O depositante se cadastra, sendo permitida a edição e consulta somente do próprio cadastro.
- Cadastro de coleção – O depositante cadastra a própria coleção, sendo permitida a edição e consulta somente da própria coleção. Um depositante pode ter várias coleções.

Caso o acesso não seja público, a primeira utilização do sistema de arquivo linguístico digital deve exigir do usuário, o uso de alguns protocolos de uso, como sua identificação e senha. Em relação ao módulo público, podemos destacar os seguintes detalhes:

- Consulta de coleção – O usuário consulta as coleções e, eventualmente, pode fazer downloads de recursos (mídias documentais) liberados pelo depositante.
- Cadastro de usuário – O usuário se cadastra, sendo permitida a edição e consulta somente do próprio cadastro. Além disso, é possível pesquisar coleções e efetuar download de recursos restritos, liberados pelo depositante/arquivista, conforme o nível de acesso.

Finalizada esta parte, ressaltamos que o módulo público (site da aplicação) deve conter opções de login (usuário e senha) para os módulos (arquivista, depositante e usuário) vistos acima. Os dois últimos cadastros são controlados pelo arquivista (administrador do ALD), mas também podem ser feitos pelos interessados por meio de preenchimento de formulário específico na tela principal do próprio

sistema.

### 3.2.4 – Diagramas de contexto

Com base nas informações referentes às necessidades ou requisitos da aplicação, passamos à construção do diagrama de contexto<sup>143</sup> para a modelagem das relações entre os componentes do sistema a fim de que tenhamos um mapa do fluxo de entrada e saída de dados.

Os diagramas abaixo representam a visão geral do sistema, no mais alto nível de abstração<sup>144</sup>, exemplificadas através da descrição gráfica do relacionamento entre as entidades externas (usuários) e o sistema no fluxo de entrada e saída de dados.

Os diagramas abaixo, referentes às figuras X, XI e XII, tratam, respectivamente, dos módulos arquivista (administrativo), depositante e usuário no sistema de arquivo linguístico digital SLMN e têm como usuários o arquivista, o depositante e o usuário-externo (internauta), conforme o seguinte delineamento:

---

<sup>143</sup> Um diagrama de contexto é um diagrama de fluxo de dados de alto nível que usa um processo único de dados para descrever o escopo e apresenta as entidades externas e os depósitos de dados que fornecem e recebem dados do sistema. Conforme o *International Institute of Business Analyses* (IIBA, 2011), os diagramas de contexto são usados em projetos compatíveis com o uso de diagramas de fluxo de dados.

<sup>144</sup> Conforme Alves (2017), devido à necessidade de compreensão de sistemas grandes e complexos, aconselha-se a sua decomposição em partes menores, momento em que é preciso o uso de alguns conceitos. Nesse caso, abstração consiste na análise sistemática do problema, desconsiderando os aspectos irrelevantes, de modo que o projetista se concentre nos pontos que realmente importam.

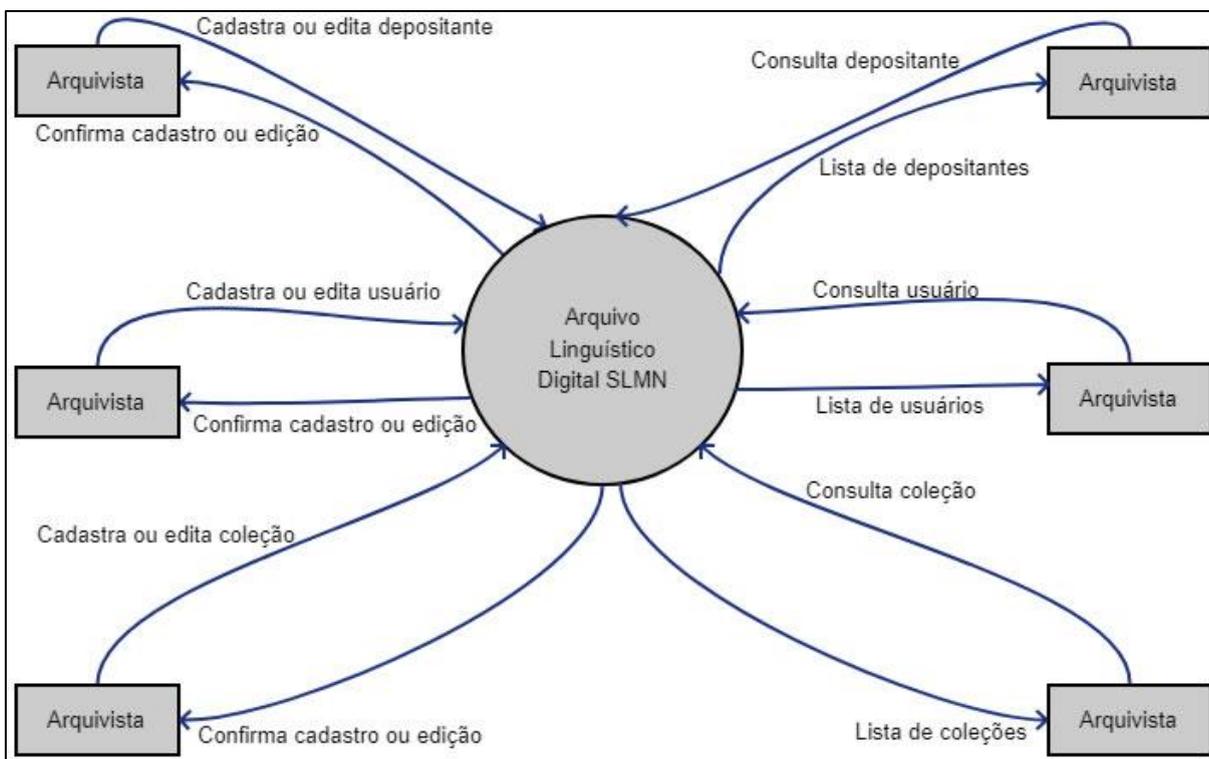


Figura X: Diagrama de contexto para as operações do arquivista<sup>145</sup>

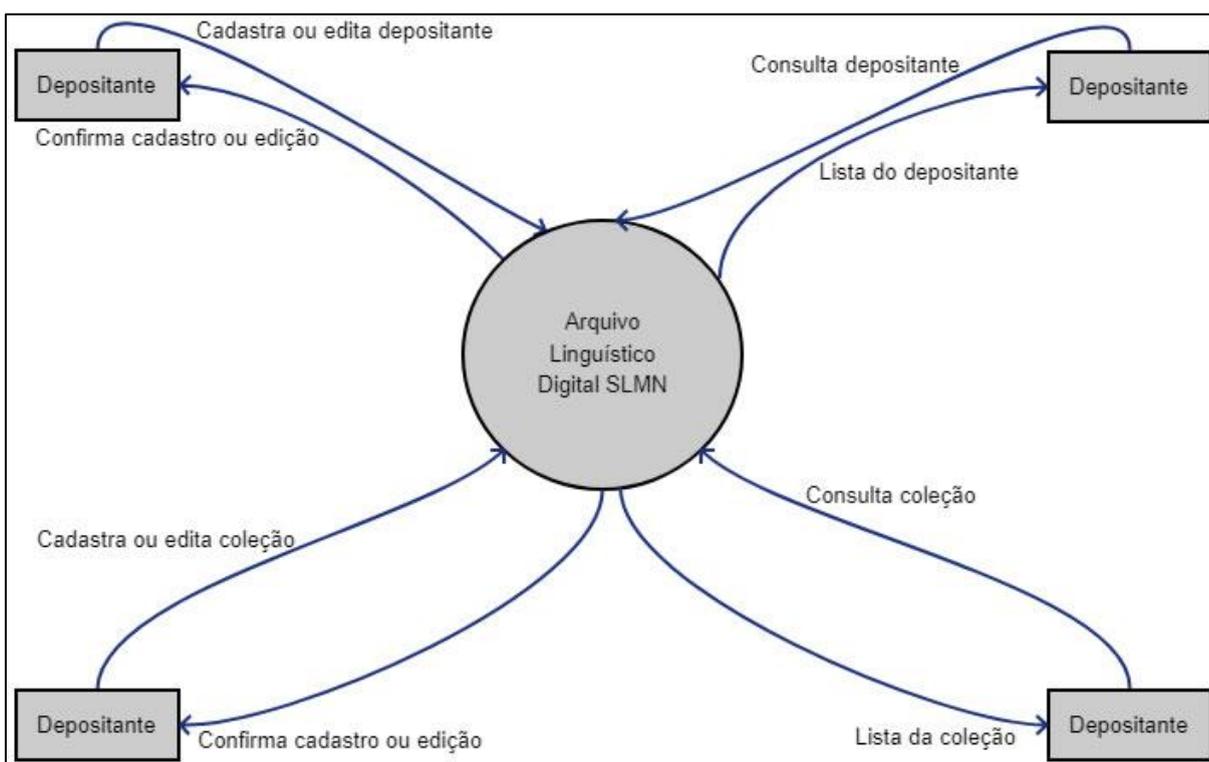


Figura XI: Diagrama de contexto para as operações dos depositantes<sup>146</sup>

<sup>145</sup> Figura retirada e adaptado de Alves (2017, p. 51).

<sup>146</sup> Figura retirada e adaptada de Alves (2017, p. 52).

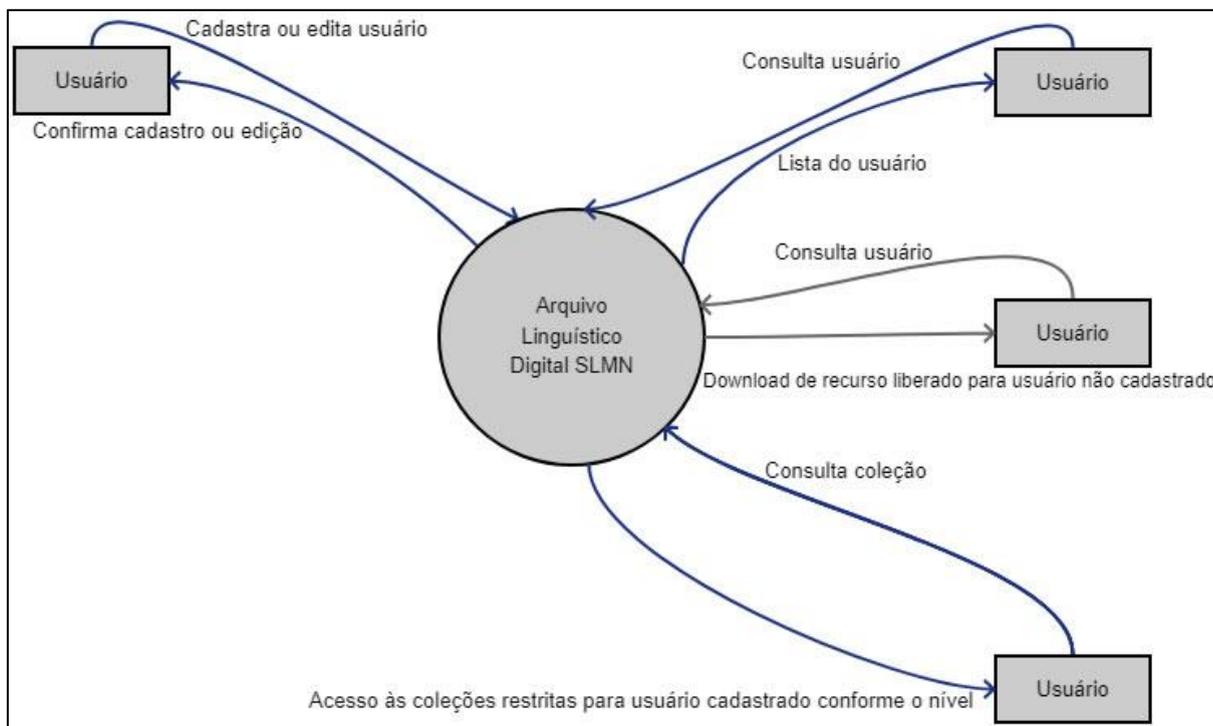


Figura XII: Diagrama de contexto para as operações dos usuários<sup>147</sup>

<sup>147</sup> Figura retirada e adaptada de Alves (2017, p. 52).

### 3.2.5 – Eventos de sistema

Os eventos de um sistema<sup>148</sup> refletem os processos operacionais que uma empresa ou instituição deseja implantar no software, ou aplicação que cogita desenvolver. No caso específico do ALD SLMN, por exemplo, o evento cadastro de coleção realizado pelo arquivista ou depositante é uma operação acompanhada pelo sistema com dois resultados possíveis: a confirmação ou rejeição do cadastro. Conforme Alves (2017), os eventos podem ser tipificados da seguinte forma:

Tipo de evento	Situação em que ocorre
Evento orientado a fluxo de dados	Todas as operações de entrada e saída são determinadas por entidades externas que comandam suas execuções
Evento temporal	O evento é executado conforme intervalo de tempo pré-determinado
Evento orientado a fluxo de controle	O evento ocorre se houver mudança de estado do atributo ou variável

Quadro IX: Classificação dos eventos<sup>149</sup>

<sup>148</sup> Cf. IIBA (2011, p. 213).

<sup>149</sup> Quadro retirado e adaptado de Alves (2017, p. 53).

Abaixo, segue o Quadro X em que alguns eventos destacados são listados conforme a situação real de uso no ALD SLMN:

Nº	Nome do evento	Tipo de Fluxo <sup>150</sup>	Estímulo	Ações	Respostas possíveis	Usuário
1	Cadastro de depositante	FD	Necessidade de registrar novo depositante	Cadastrar depositante	Cadastro feito com sucesso. Cadastro já existe.	Arquivista
2	Consulta coleção	FD	Usuários e/ou pesquisadores pedem informações sobre coleções específicas acesso restrito, etc.	Consultar coleções de acesso restrito	Lista das coleções restritas e contato do depositante para possível liberação ao acesso	Arquivista
3	Cadastro de coleção	FD	Necessidade do depositante de registrar coleção	Cadastrar coleção	Cadastro feito com sucesso. Cadastro já existe.	Depositante
4	Pesquisa de coleção via web	FD	Interesse do usuário externo (internauta) em pesquisar coleção	Pesquisar coleção no ALD SLMN	Lista das coleções abertas e possível download de recurso. O acesso a coleções restritas exige login e senha	Usuário externo (internauta)

Quadro X: Lista dos principais eventos executados<sup>151</sup>

### 3.2.6 – Estrutura de dados

Através dos diagramas vistos, anteriormente, percebemos que a representação da utilização do sistema pelos usuários (arquivista, depositante e internautas) indica o uso intenso da base de dados onde estão armazenados os dados do arquivo linguístico SLMN. Diante disso, ressaltamos essa modelagem comportamental<sup>152</sup>

<sup>150</sup> Significado da legenda correspondente à coluna “Tipo de fluxo”: fluxo de dados (FD), fluxo temporal (FT), fluxo de controle (FC).

<sup>151</sup> Quadro retirado e adaptado de Alves (2017, p. 53).

<sup>152</sup> Conforme Pompilho (1995), a modelagem comportamental descreve como o sistema reage, internamente, aos estímulos externos. Na prática, seriam as ações que o sistema executa ao responder aos eventos previstos no modelo ambiental. Compõem o modelo comportamental os Diagramas de Fluxo de Dados (DFD), os Dicionários de Dados e os Diagramas Entidade-Relacionamento (DER).

como etapa importante na compilação das informações essenciais para a composição da estrutura do banco de dados que aglutinará, prioritariamente, dados de documentação linguística.

A estruturação da base de dados SLMN origina-se da análise cuidadosa das principais características e documentos requisitados pelos ALD, DOBES, DMA e AILLA, principalmente deste último, que subsidiaram os formulários de cadastro de seus respectivos sistemas linguísticos (ver capítulo 2 desta dissertação). O resultado desses estudos contribuiu para que compuséssemos a estrutura interna de nosso arquivo linguístico digital a fim de atender as demandas básicas de nossos pesquisadores ao considerar a capacidade do ALD em armazenar a variedade de corpus documentais digitais criados durante a pesquisa de campo. Assim, concluímos que a base de dados em tela deve conter as seguintes especificações e entidades<sup>153</sup>:

- Cadastro de usuário do sistema
- Cadastro de coleção linguística
- Cadastro do depósito na coleção
- Cadastro da família linguística
- Cadastro da língua pesquisada

---

<sup>153</sup> Entidades são objetos existentes no mundo real que possuem diferentes identidades e significados. No ambiente de banco de dados, são a representação de uma classe de dados no negócio, ou seja, um conjunto de informações de mesma característica cujas instâncias (ocorrências) identificam esses dados (MACHADO, 2008).

As cinco entidades que compõem a base de dados do ALD SLMN podem ter suas estruturas internas representadas conforme a sequência dos quadros (XI a XV) que se segue:

Nome do atributo	Descrição	Tipo de chave	Tipo de dado	Tamanho	Formatação	Requerido
Codigo_usuario	Código de identificação do usuário	Primária	Numérico			Sim
Nome_usuario	Nome do usuário		Alfanumérico	20		Sim
Sobrenome	Sobrenome do usuário		Alfanumérico	20		Sim
e-mail	Email do usuário		Alfanumérico	30		Sim
Instituicao	Instituição do usuário		Alfanumérico	40		Sim
Pais	País de origem do usuário		Alfanumérico	25		Sim
Senha_acesso	Senha de acesso ao sistema		Alfanumérico	10		Sim
Perfil_acesso	Identifica o usuário	Estrangeira	Numérico		1-arquivista 2-depositante 3-internauta	Sim
Nível_acesso	Código do nível de acesso		Numérico		1 (mais baixo) 2 3 4 (mais alto)	Sim

Quadro XI: Entidade usuário do sistema<sup>154</sup>

<sup>154</sup> Quadro retirado e adaptado de Alves (2017, p. 54).

Nome do atributo	Descrição	Tipo de chave	Tipo de dado	Tamanho	Formatação	Requerido
Codigo_colecao	Código de identificação da coleção	Primária	Numérico			Sim
Titulo_portugues	Título da coleção em português		Alfanumérico	50		Sim
Titulo_ingles	Título da coleção em inglês		Alfanumérico	50		Sim
Lingua_pesquisada	Língua cadastrada na coleção	Estrangeira	Numérico			Sim
Depositante	Depositante da coleção	Estrangeira	Numérico			Sim
Local_projeto_colecao	Local onde projeto da coleção se desenvolve	Estrangeira	Numérico			Sim
Descricao_colecao	Descrição das características da coleção		Alfanumérico	100		Sim
Referenciais_teoricos	Lista dos referenciais teóricos usados		Alfanumérico	100		Sim

Quadro XII: Entidade coleção linguística<sup>155</sup>

<sup>155</sup> Quadro retirado e adaptado de Alves (2017, p. 55).

Nome do atributo	Descrição	Tipo de chave	Tipo de dado	Tamanho	Formatação	Requerido
Codigo_deposito	Código de identificação do depósito	Primária	Numérico			Sim
Codigo_colecao	Código da coleção relacionada ao depósito	Estrangeira	Numérico			Sim
Codigo_depositante	Código do depositante responsável pela coleção	Estrangeira	Numérico			Sim
Codigo_ISO_Lingua	Código da língua de acordo com padrão ISSO	Estrangeira	Numérico			Sim
Codigo_ETH_lingua	Código da língua de acordo com padrão Ethnologue	Estrangeira	Numérico			Sim
Num_deposito	Número sequencial do depósito feito na coleção		Numérico			Sim
Num_recurso	Número sequencial do novo recurso depositado na coleção		Numérico			Sim
Num_itens	Número incremental de itens associados ao recurso		Numérico			Sim
Codigo_extensao	Descreve a extensão da mídia associada ao item do recurso	Estrangeira	Numérico		.txt; .doc; .docx; pdf; .mp3; .mp4; .wav;.rtf; .xls; xml; etc.	Sim
Deposito_midia	Armazena a string com o caminho onde a mídia está armazenada no disco		Alfanumérico			Sim

Quadro XIII: Entidade depósito na coleção<sup>156</sup><sup>156</sup> Quadro retirado e adaptado de Alves (2017, p. 56).

Nome do atributo	Descrição	Tipo de chave	Tipo de dado	Tamanho	Formatação	Requerido
Codigo_familia	Código de identificação da família linguística	Primária	Numérico			Sim
Codigo_ETH_familia	Código da família de acordo com padrão Ethnologue	Estrangeira	Numérico		<i>Display o nome da família</i>	Sim
Numero_de_linguas	Número de línguas que compõem a família linguística		Numérico			Sim

Quadro XIV: Entidade família linguística<sup>157</sup>

<sup>157</sup> Quadro retirado e adaptado de Alves (2017, p. 57).

Nome do atributo	Descrição	Tipo de chave	Tipo de dado	Tamanho	Formatação	Requerido
Codigo_lingua	Código de identificação da língua pesquisada	Primária	Numérico			Sim
Codigo_ETH_lingua	Código da língua de acordo com padrão Ethnologue	Estrangeira	Numérico		<i>Display o nome da língua</i>	Sim
Codigo_ISO_Lingua	Código da língua de acordo com padrão ISSO	Estrangeira	Numérico			Sim
Codigo_ETH_familia	Código da família de acordo com padrão Ethnologue	Estrangeira	Numérico		Display do Nome da família	Sim
Locais_lingua_falada	Descrição dos locais (estados) onde a língua é falada		Alfanumérico	200		Sim
Pais_lingua_falada	Descrição do(s) país(es) onde a língua é falada		Alfanumérico	200		
Link_lingua_ETH	Link da língua no site do Ethnologue		Alfanumérico	100		Sim

Quadro XV: Entidade língua pesquisada<sup>158</sup>

<sup>158</sup> Quadro retirado e adaptado de Alves (2017, p. 57).

A figura abaixo mostra a representação, simplificada, do sistema de arquivo linguístico digital SLMN web através da conexão relacional entre as principais entidades que compõem seu banco de dados:

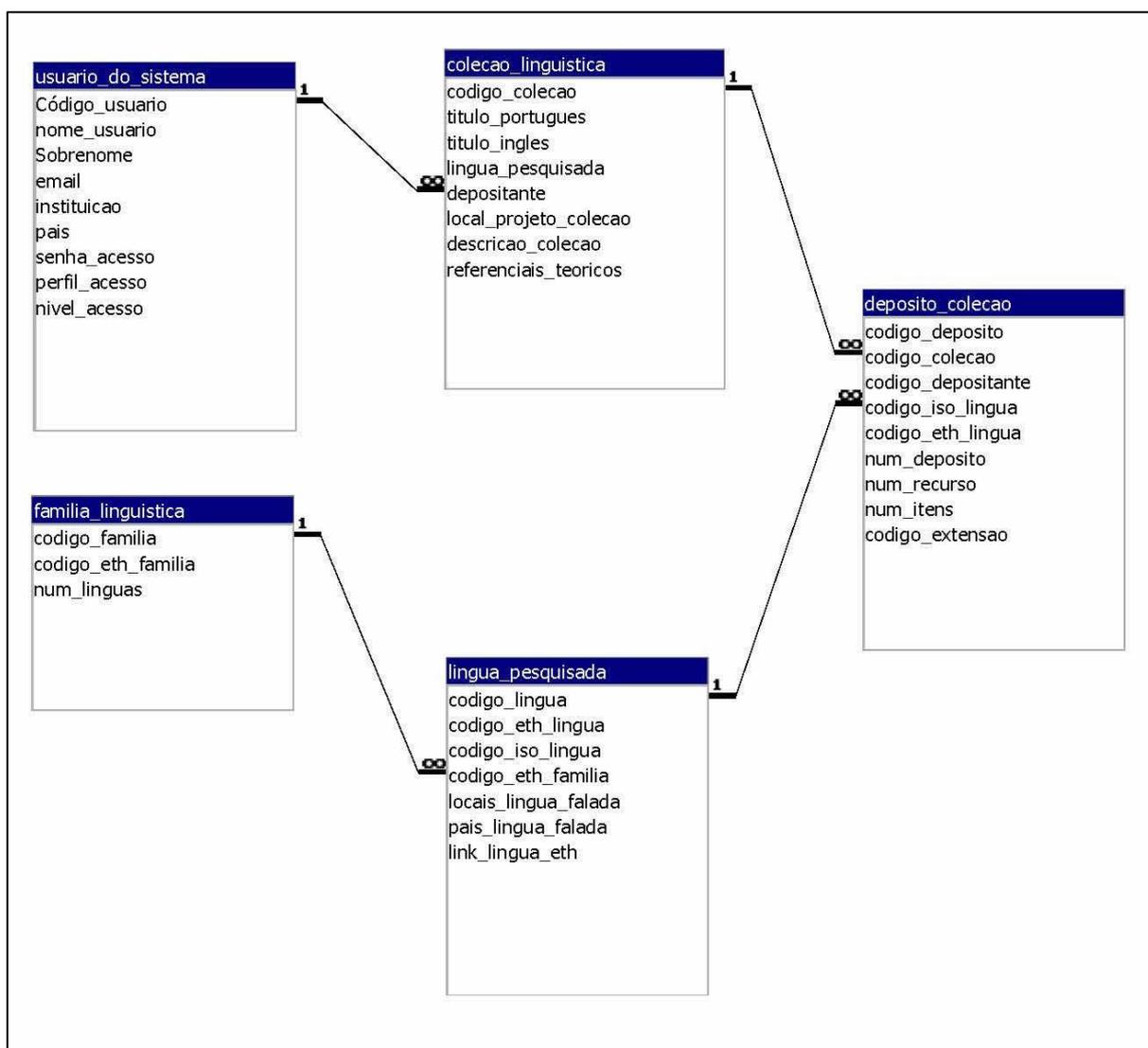


Figura XIII: Estrutura entidade-relacionamento (ER) no banco de dados<sup>159</sup>

<sup>159</sup> A abordagem entidade-relacionamento criada em 1976 por Peter Chen é a técnica de modelagem de dados mais difundida para a especificação de requisitos e modelagem conceitual que objetiva a projeção de sistemas conforme a especificação levantada pelo analista de sistemas junto às partes interessadas (ver HEUSER, 2004; MACHADO, 2008; TEOREY, 2014).

Sobre as relações estabelecidas entre as entidades acima, depreende-se o seguinte:

- Na relação entre as entidades coleção linguística e usuário do sistema: a entidade coleção linguística pode ser utilizada por vários usuários (arquivista, depositante e internauta).
- Na relação entre as entidades coleção linguística e depósito coleção: a entidade coleção linguística pode ter vários depósitos efetuados pelo arquivista ou depositante.
- Na relação entre as entidades depósito, coleção e língua pesquisada: a mesma língua pode ser objeto de estudo de pesquisadores diferentes e fazer parte de outras coleções.
- Na relação entre família linguística e língua pesquisada: a mesma família linguística pode ter várias línguas pesquisadas.

### 3.2.7 – Interface sistema-usuário

Chegamos ao final deste projeto do sistema ALD SLMN visando fundamentar os processos de modelagem do desenvolvimento das interfaces humano-computador, descrevendo-os desde o processo de criação dos diagramas da estrutura de cada um dos módulos até a definição dos layouts das telas da aplicação web. As técnicas de elaboração das telas de interface, apresentadas mais adiante, neste tópico, têm relação direta com o novo ramo da computação conhecido como “Interface humano-computador”, em que o objetivo é estudar a interação entre os usuários e os sistemas computacionais a fim de simplificar sua utilização, tornando-os mais produtivos. Em linha com essas diretrizes, a finalidade de cada aplicação exige uma interface específica que atenda as exigências do consumidor de software, como, por exemplo, a facilidade de uso (BEVAN, 1999); neste caso, a equipe de desenvolvimento define os protótipos de layout de interface com o futuro usuário.

Com base nas discussões feitas até aqui, apresentamos, abaixo, algumas figuras que representam, hierarquicamente, os módulos que compõem o sistema ALD SLMN. Começamos pelo módulo arquivista, tendo em vista sua maior frequência de uso, além da relação de cada uma das páginas web entre si, e de como cada uma delas ativam umas às outras.

### 3.2.7.1 – Estrutura hierárquica dos módulos

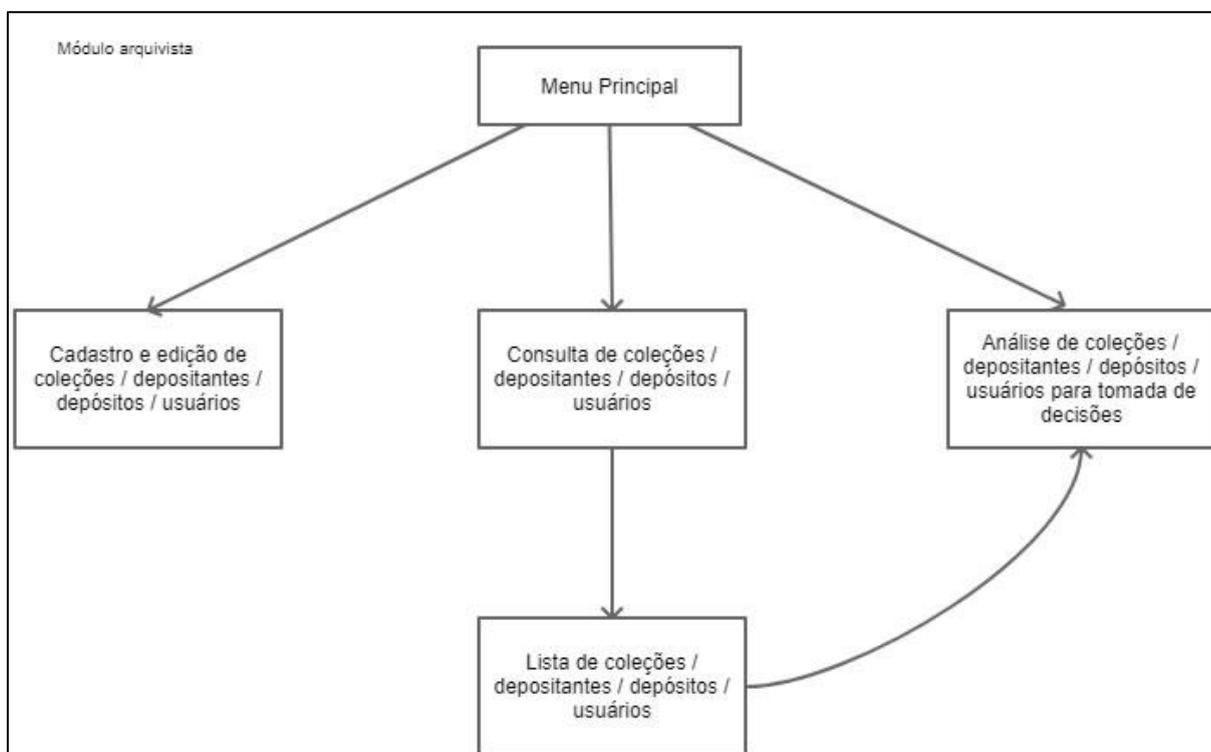


Figura XIV: Estrutura hierárquica do módulo arquivista<sup>160</sup>

No módulo arquivista, a página inicial mostra: 1) a consulta à coleção, ao depositante, ao depósito ou ao usuário; 2) o resultado da consulta lista as coleções, os depositantes, os depósitos ou usuários existentes no ALD; 3) caso a coleção, o depositante, o depósito ou o usuário não esteja cadastrado, o arquivista poderá fazê-lo na rotina de cadastro apropriada para essa tarefa.

<sup>160</sup> Figura retirada e adaptada de Alves (2017, p. 66-67).

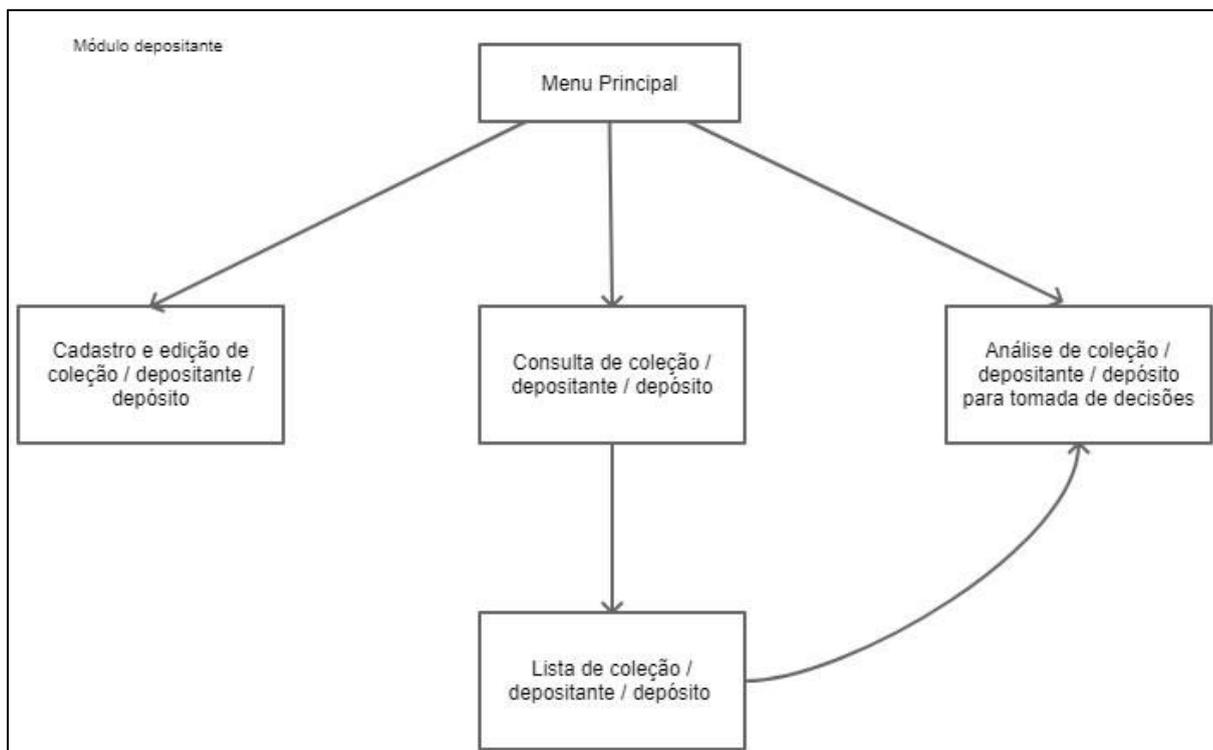


Figura XV: Estrutura hierárquica do módulo depositante<sup>161</sup>

No módulo depositante, a página inicial mostra: 1) a consulta à coleção, ao depositante ou ao depósito; 2) o resultado da consulta lista o depositante, sua coleção e depósitos; 3) caso a coleção, o depositante, ou o depósito não esteja cadastrado, o depositante poderá fazê-lo na rotina de cadastro apropriada para essa tarefa.

<sup>161</sup> Figura retirada e adaptada de Alves (2017, p. 66-67).

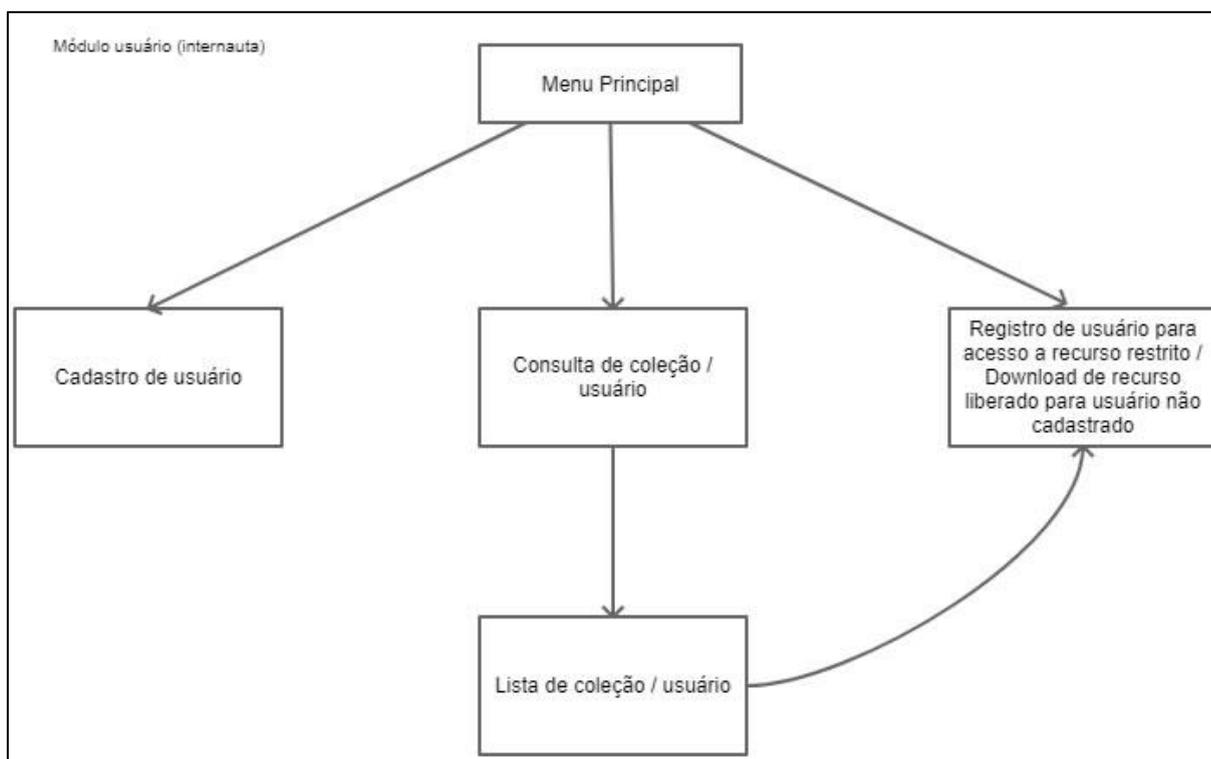


Figura XVI: Estrutura hierárquica do módulo usuário<sup>162</sup>

No módulo usuário (internauta), a página inicial mostra: 1) a consulta à coleção ou ao usuário; 2) o resultado da consulta lista as coleções existentes no ALD ou o próprio usuário cadastrado; 3) caso o usuário deseje baixar um recurso restrito, será exigido o cadastro, por outro lado, se o download do recurso estiver liberado, o cadastro não será exigido.

<sup>162</sup> Figura retirada e adaptada de Alves (2017, p. 66-67).

### 3.2.7.2 – Modelos de tela na interface sistema-usuário

Encerrada a apresentação de todas as partes que estruturam o projeto ALD SLMN, com objetivo ilustrativo e complementar a todos os elementos conceituais expostos nesta seção, destacamos, abaixo, a prévia de cinco telas da interface<sup>163</sup> sistema-usuário: página principal (módulo público), login de usuário (módulo público), cadastro de usuário (módulos arquivista, depositante e internauta), cadastro de coleção (módulos arquivista e depositante) e depósito na coleção (módulos arquivista e depositante).

Todas as páginas do sistema ALD web possuem o mesmo layout a fim de padronizar a identidade gráfica do site. Na página principal (ver Figura XVII), a barra superior é formada pela área do cabeçalho que reserva a imagem de fundo, o nome do usuário logado, o título do site e seu logotipo. No rodapé, é possível mostrar os direitos de copyright, o ano de criação do site, entre outras informações.

Logo abaixo, da barra superior, estão localizadas as opções do menu principal (Cadastro, Consulta, Depósito, Relatório e Login). A parte central da tela divide-se em duas colunas. O canto esquerdo, superior, foi reservado para informes sobre as coleções disponíveis para consulta; e o esquerdo, inferior, foi reservado para a publicação de notícias relevantes sobre as pesquisas na área de documentação. A coluna à direita foi destacada para informar aos usuários sobre os objetivos e orientações gerais de uso do ALD SLMN.

Para todas as atividades que envolvem formulários, a área centro-esquerda é utilizada para cadastro e consulta de dados armazenados no arquivo. Com esse fito, exporemos uma fração, a título de exemplo, das telas, com um pequeno resumo de suas funcionalidades, lembrando que é através da página principal que o usuário acessa o menu com todas as opções de uso do sistema.

---

<sup>163</sup> Para as telas de interface das aplicações que funcionarão na Internet, Montero et al. (2002) propõe os modelos para projetos web agrupados em três categorias: website, páginas web e ornamentais.

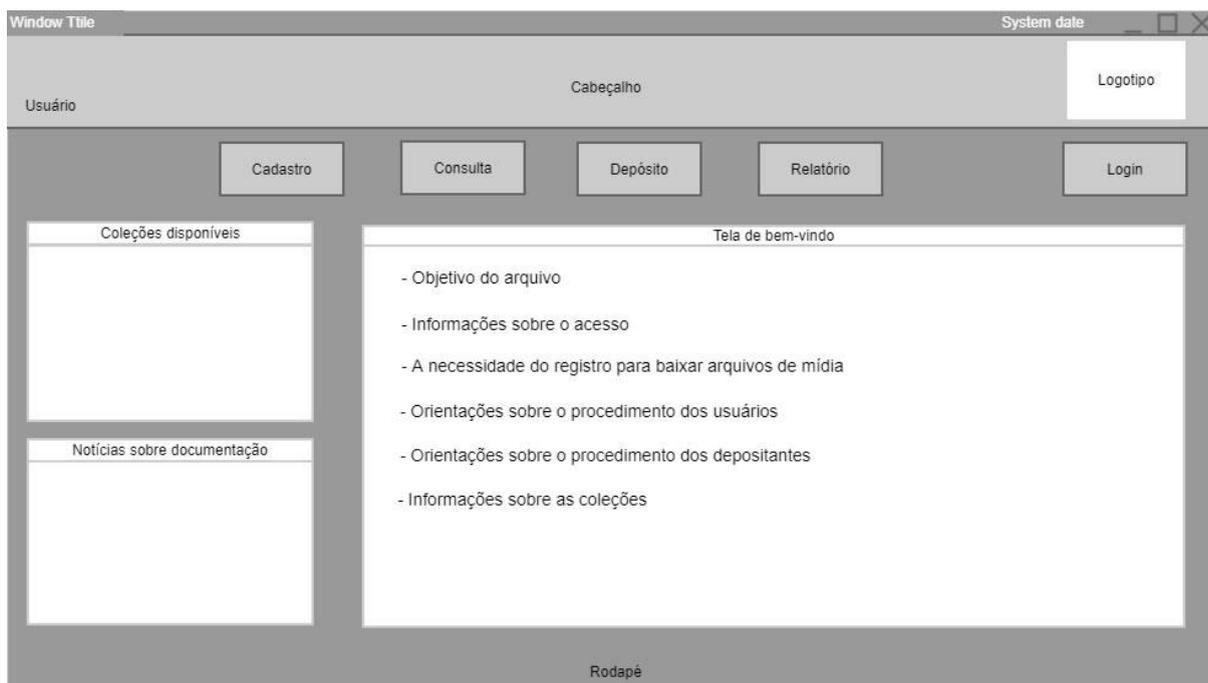


Figura XVII: Tela de interface da página principal<sup>164</sup>

A tela de login (módulo público), abaixo (ver Figura XVIII), permite ao usuário cadastrado fazer login, usando as credenciais (login e senha), como arquivista, depositante ou usuário externo (internauta) ao pressionar o botão “Entrar”. Através das credenciais de acesso utilizadas, o sistema controla as permissões de uso do ALD, ou seja, o escopo de cada um dos perfis. O mesmo procedimento também é gerenciado pelo arquivista a partir de seu próprio módulo. Caso o usuário não possua as credenciais necessárias para usar o ALD, é possível o preenchimento de formulário com essa finalidade ao clicar no botão “Cadastrar”.

<sup>164</sup> Figura retirada e adaptada de Alves (2017, p. 69-75).

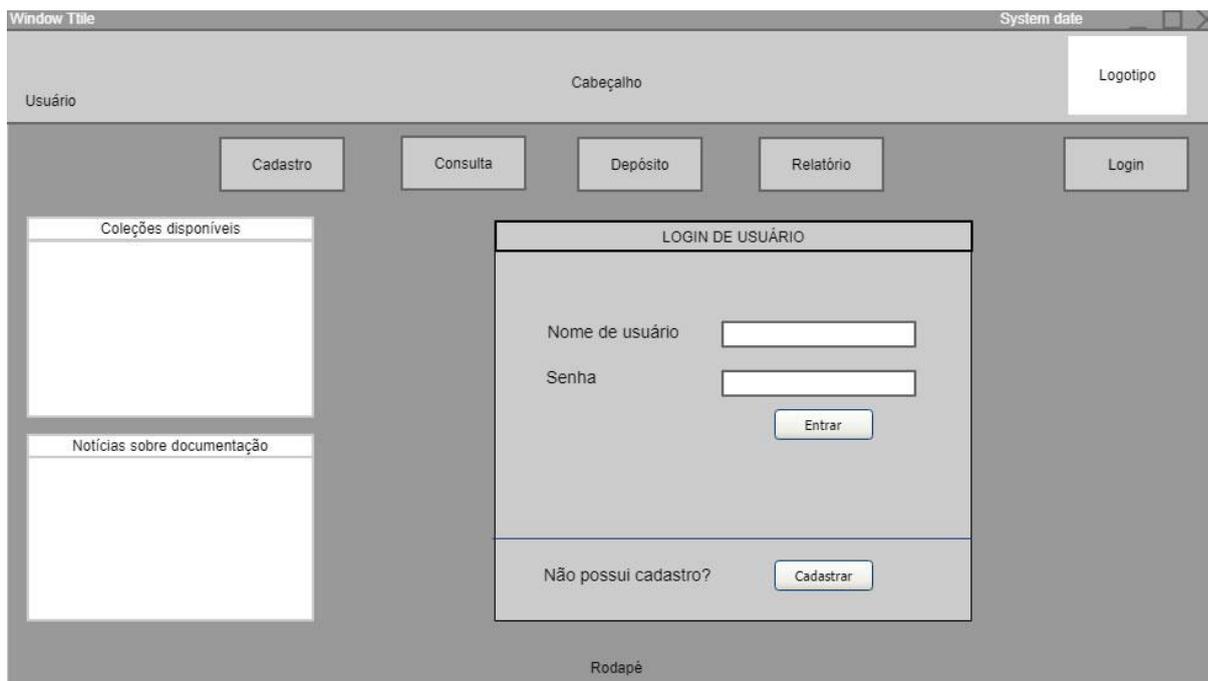


Figura XVIII: Tela de interface do login de usuário<sup>165</sup>

A tela de cadastro de usuário (módulos arquivista, depositante e usuário externo) possibilita ao depositante e ao internauta efetuarem o cadastro básico (ver Figura XIX) ao clicar no botão “Cadastrar”, unilateralmente, mas somente ao arquivista é permitido controlar seus cadastros no que tange à definição dos perfis e níveis de acesso. Quando é o próprio usuário que faz o cadastro, não há acesso ao perfil e ao nível de acesso, pois somente o arquivista manipula essas opções. Nesse caso, o nível de acesso, por default, é 1.

<sup>165</sup> Figura retirada e adaptada de Alves (2017, p. 66-67).

Figura XIX: Tela de interface do cadastro de usuário<sup>166</sup>

A tela de cadastro de coleção (módulos arquivista e depositante) permite que arquivista e depositante cadastrem novas coleções linguísticas no ALD (ver Figura XX). No caso do depositante, em particular, é necessário que seu nível de acesso lhe conceda essa permissão. No ALD, o depositante pode possuir várias coleções e essas coleções podem receber depósitos sucessivos em seu acervo. Após o preenchimento do formulário, a operação é finalizada pressionando o botão “Cadastrar”.

<sup>166</sup> Figura retirada e adaptada de Alves (2017, p. 66-67).

Figura XX: Tela de interface do cadastro de coleção<sup>167</sup>

A tela depósito na coleção (módulos arquivista e depositante) permite ao arquivista e depositante efetuarem novos depósitos na coleção linguística (ver Figura XXI). Por padrão, o depositante está habilitado a fazer depósito na própria coleção, ou coleções, se houver mais de uma. Na impossibilidade de o depositante fazer a operação, o arquivista (administrador do ALD) pode executá-la em seu lugar. Após o preenchimento de todos os campos do formulário, a aplicação é finalizada ao pressionar o botão “Depositar”.

<sup>167</sup> Figura retirada e adaptada de Alves (2017, p. 66-67).

Figura XXI: Tela de interface do depósito na coleção<sup>168</sup>

Chegamos a este ponto, reforçando duas questões importantes nesta proposta de elaboração de ALD: o prévio levantamento completo dos requisitos necessários para a elaboração do modelo conceitual e a recomendação de se publicar o futuro arquivo linguístico digital no ambiente virtual, tecnicamente, mais seguro, em termos de cibersegurança para esse tipo de BD, como a Intranet. Esse espaço virtual, somente, não garante toda a segurança, é necessário também que o levantamento de requisitos transposto para o modelo conceitual seja aplicado<sup>169</sup>. Ressaltamos que tais sugestões não têm como esteio ortodoxias inflexíveis, mas observações críticas feitas a partir de nossas análises (ver item 3.1: “Análise de resultados”), sujeitando-se, portanto, a revisões de caráter científico que podem ser feitas futuramente.

Sinalizando o término dessa etapa da apresentação do projeto, indicamos os elementos mais concretos sobre a confecção do sistema de arquivo linguístico. O início desta seção aponta algumas sugestões em termos de software que podem ser utilizados pelo desenvolvedor a ser designado para a materialização da aplicação apresentada no projeto básico de ALD.

<sup>168</sup> Figura retirada e adaptada de Alves (2017, p. 66-67).

<sup>169</sup> Conforme Blackmore (2005, p. 6): “Uma Intranet é uma rede corporativa privada que usa produtos e tecnologias da Internet. O acesso a uma Intranet é controlado pela organização que a estabeleceu e geralmente é restrito apenas aos funcionários. Ocasionalmente, no entanto, fornecedores e clientes também podem ter acesso a partes dele”. Para questões relacionadas às políticas de segurança na Intranet, ver Colby et al. (2003, p. 207-208).

Para além das proposições de carácter técnico, sugerimos, caso a coordenação do Setor de Linguística do Museu Nacional-UFRJ encontre dificuldades de ordem económico-financeira para o desenvolvimento da aplicação no espaço privado, a possibilidade de recorrer-se ao conhecimento potencial dos formandos do curso de Ciências da Computação<sup>170</sup> da UFRJ na fase do estágio. Este curso prevê que o corpo discente faça essa atividade em seu projeto pedagógico para aprovação e obtenção do diploma (§1º do art. 2º da Lei nº 11.788/2008)<sup>171</sup> no âmbito de empresas privadas ou de instituições públicas.

Assim, cremos que na impossibilidade de se contratar empresa de projeto e desenvolvimento de sistemas, o estagiário da UFRJ<sup>172</sup>, como futuro profissional, acompanhado do orientador, pode representar a possibilidade da construção de um arquivo linguístico que siga os mesmos parâmetros qualitativos daqueles analisados nesta dissertação. Esse desafio seria ainda mais produtivo para o SLMN se todas as etapas desse trabalho fossem acompanhadas por um técnico administrativo que tenha afinidade com a informática, esteja disposto a receber treinamento em desenvolvimento de sistemas e, futuramente, possa ser aquele que fará a intermediação junto ao curso de Ciências da Computação-UFRJ frente às necessidades de atualização do software linguístico conforme seu ciclo de vida.

Finalizando esta seção, chamamos a atenção de que todas as etapas previstas para a construção do arquivo linguístico digital, direcionado ao Setor de Linguística do Museu Nacional, foram contempladas. Com isso, acreditamos firmemente que os fundamentos teóricos e metodológicos essenciais para a criação de uma aplicação web, de médio porte, foram devidamente explorados. Apoiados pelos estudos feitos nesta dissertação, a proposta é contribuir para possibilitar o desenvolvimento real de uma aplicação que atenda às necessidades de pesquisadores, professores, estudantes e comunidades indígenas. Nesse caso, o ALD proposto pode auxiliar armazenando, organizando e publicizando o corpus documental resultante das pesquisas em andamento, das pesquisas recentes e das pesquisas antigas, caso estejam dispersas e/ou sob o risco iminente de se perderem. Assim, estamos

---

<sup>170</sup> Cf. Pro-Reitoria de Graduação – PR1. Disponível em: <<https://xn--graduao-2wa9a.ufrj.br/index.php/cursos-de-graduao-mainmenu-124?id=106:cincias-da-computao&catid=158:c>>. Acesso em: 5 de nov. 2022.

<sup>171</sup> Cf. Cartilha sobre a Lei do Estágio. Disponível: <[https://dcc.ufrj.br/attachments/article/66/cartilha\\_Lei\\_Estagio.pdf](https://dcc.ufrj.br/attachments/article/66/cartilha_Lei_Estagio.pdf)>. Acesso em: 6 nov. 2022.

<sup>172</sup> Cf. Normas – Programa de estágio. Disponível em: <https://dcc.ufrj.br/ensino/graduacao/programa-de-estagio-normas>. Acesso em: 7 nov. 2022.

convencidos de que um banco de dados com esses recursos tem grande possibilidade de tornar sustentáveis as pesquisas produzidas no SLMN, possibilitando a formação de um acervo que pode ser referência nos estudos linguísticos direcionados à documentação das línguas indígenas ameaçadas.

### 3.3 – Considerações finais

Como qualquer trabalho científico que não esgota todas as reflexões possíveis sobre o referencial acadêmico relacionado a seu objeto, destacamos algumas limitações deixadas para investigações futuras. Uma delas está relacionada à escassez de produções acadêmicas textuais que descrevam ou analisem os programas de documentação e seus respectivos arquivos linguísticos digitais.

Outra dificuldade foi provocada pela impossibilidade de desenvolvimento do estudo de caso, por meio de atividade de campo, prevista em nosso primeiro projeto de pesquisa, devido ao surto de COVID-19 que atingiu duramente a população mundial, principalmente os povos mais fragilizados, entre os quais os indígenas, fato que afetou, sensivelmente, nossa proposta metodológica, limitando nossas investigações à pesquisa bibliográfica.

Como objeto de nossos estudos, elegemos os PDL (DOBES, DMA e AILLA), com ênfase nas abordagens que se debruçam sobre as línguas indígenas ameaçadas de extinção. Como vimos, ao observar o processo histórico que determinou o surgimento e consolidação desses esforços acadêmicos, devem ser considerados seu envolvimento prioritário na situação crítica enfrentada pelos falantes associados ao *boom* de investimentos realizados na primeira década do século XX e, no período seguinte, por uma nova realidade marcada pelo profissionalismo e comprometimento dos especialistas em documentação linguística.

Em linha com essa discussão, os elementos metodológicos seguintes foram fundamentais para que pudéssemos formular conclusões sobre nosso objeto: pesquisa aplicada, pesquisa exploratória, pesquisa quali-quantitativa, método hipotético-dedutivo, pesquisa bibliográfica e pesquisa documental.

Em relação à resposta ao problema (teórico) desta dissertação, podemos, em suma, dizer que a submissão dos programas aos três princípios estruturantes de um PDL (1 - o enfrentamento à crise da diversidade linguística; 2 - o aproveitamento dos recursos financeiros destinados à pesquisa documental; e 3 - a verificabilidade dos dados primários por pares.), definidos por Himmelmann (2007, p. 46–47), voltados ao combate à crise da diversidade linguística, apresentou os seguintes resultados: DOBES e AILLA, cumprimento pleno; o DMA foi exceção (princípio da verificabilidade). Entretanto, esse resultado não invalida as outras iniciativas deste PDL. Assim, podemos afirmar que o impacto dos três programas de documentação

no enfrentamento da ameaça à diversidade linguística, com destaque para as línguas indígenas, é altamente positivo e produtivo. Para informações completas sobre o problema (teórico) da pesquisa (capítulo 1), e as análises das respostas dadas pelos PDL ver capítulo 2, seção 2.5.

Em resposta ao objetivo geral da pesquisa (ver nesta dissertação, o capítulo 1), verificamos que os programas de documentação linguística analisados impactaram, conjuntamente, sobre a questão da ameaça às línguas indígenas a partir da elevação do volume de investimentos que permitiram: o aumento da produção acadêmica, o incremento do volume de recursos investidos em pesquisa, a promoção de diversos eventos acadêmicos (simpósios, workshops, etc.), a criação de novas técnicas de arquivamento (digital), o envolvimento de variadas instituições em novos projetos e a criação frequente de novas tecnologias direcionadas à documentação, entre outras iniciativas.

Em relação aos três objetivos específicos definidos (ver nesta dissertação, capítulo 1), destacamos os seguintes resultados:

- Entre as características comuns destacadas entre os PDL (DOBES, DMA e AILLA), direcionadas à documentação das línguas indígenas ameaçadas, estão a abrangência global, as pesquisas financiadas por agências governamentais e privadas, o alto nível tecnológico, a possibilidade de gravação de mídia analógica e/ou digital, a utilização de arquivo linguístico digital e a fragilidade dos ALD instalados no ambiente web.
- Identificamos que todos programas apresentam arquivos linguísticos digitais com problemas nos protocolos de acesso dos usuários e, conseqüentemente, algum risco para a integridade dos materiais arquivados. A iniciativa DOBES de permitir que outras instituições utilizem sua tecnologia de BD lhe garante o armazenamento grande volume de materiais linguísticos de pesquisadores na matriz europeia ou nos países conveniados.
- Entre os ALD analisados, aquele que apresenta os melhores recursos, pelos fatores citados, é o DOBES. Porém, aquele que tem as características mais adequadas (tecnologia GPL com baixa curva de aprendizagem) à realidade de linguistas, donos de acervos particulares e instituições de pequeno e médio porte interessadas nesse tipo de solução, é o AILLA.

A partir daquilo que vimos e analisamos nos capítulos 1 e 2, podemos concluir

que o impacto dos PDL (DOBES, DMA e AILLA) sobre as línguas indígenas ameaçadas é positivo, satisfatório e extremamente importante nas tentativas de se evitar uma catástrofe linguística e cultural irreparável para a humanidade. O que explica a relevância de especialistas e setores da sociedade civil mobilizados em torno da causa.

Portanto, podemos afirmar que até os dias atuais, os programas de documentação linguística analisados fizeram muito, porém não tudo que é possível e necessário em prol dos povos indígenas de línguas ameaçadas, dos indígenas remanescentes de povos cuja língua foi apagada e desejam restaurá-la e daqueles grupos indígenas que querem documentar as suas línguas, mesmo que estas estejam vivas, pela manutenção da transmissão intergeracional e pelo número de falantes, precavendo-se de um futuro incerto.

Nos aproximamos do encerramento das considerações finais, afirmando que esta dissertação representa uma das atividades de pesquisa, ao nível bibliográfico, mais abrangentes em torno das questões relacionadas a programas de documentação, envolvendo discussões teóricas, metodológicas e tecnológicas acerca da proposta de se fazer documentação linguística direcionada às línguas indígenas brasileiras.

O material elaborado para cumprir a etapa de conclusão do Curso de Mestrado Profissional em Linguística e Línguas Indígenas do PROFLLIND/MN/UFRJ, possui características de um *handbook* (manual) que pode ser útil para os professores do referido curso, professores externos e estudantes interessados em ingressar no universo da documentação linguística, pois como dissemos, abrange praticamente todos os elementos necessários para a sua aplicação no mundo real da pesquisa de campo, com destaque para a seção 3.2 que possui uma proposta básica de projeto de ALD.

Creio que o conjunto das abordagens desenvolvidas neste trabalho delineiam novos caminhos para os especialistas que desejam documentar conforme os padrões estabelecidos pelas melhores práticas. Como expliquei, uma das etapas da pesquisa foi suprimida, por isso recomendo que futuras pesquisas sobre esse tema ou assemelhadas, não prescindam (a não ser em casos excepcionais, como os pandêmicos) de fazer um estudo de caso associado às atividades no campo. Os resultados dos dados coletados no campo contribuirão para que o pesquisador coteje o levantamento bibliográfico com o empirismo que, realisticamente, indicará aquilo que

é, verdadeiramente, viável em termos teóricos e metodológicos, além de habilitá-lo a avaliar quais são as condutas éticas a adotar no *habitat* da comunidade de fala.

Para finalizar, considero que esta tenha sido uma das atividades de pesquisa mais desafiadoras por mim enfrentada. O longo período de sua elaboração se conformou em meio a uma das maiores adversidades de saúde já enfrentadas pela humanidade nos últimos cem anos, com números revistos que indicam, até a data que escrevo, uma perda de quase 20 milhões de pessoas no mundo.

Em meio a essa tragédia, precisei conciliar minhas atividades profissionais com as acadêmicas. Apesar das dificuldades, considero que tenha sido uma experiência amplamente positiva em termos de acesso à produção de conhecimento científico. No primeiro momento, o contato com o estado da arte do tema abordado, e, posteriormente, o mergulho nos mais profundos estudos acadêmicos, de modo que as reflexões expostas possibilitaram a produção do material acadêmico-profissional sobre o qual me debruço.

Esta obra, a partir de agora, tem vários donos e coautores, entre os quais aqueles que desejam, em algum momento, criar um programa de documentação linguística ou, simplesmente, queiram armazenar corpus documentais resultantes de pesquisas de campo - geralmente guardados em locais esquecidos e sob o risco de perda -, em um acessível e seguro arquivo linguístico digital.

## REFERÊNCIAS

- ABREU, R. N. Estatutos jurídicos e processos de nacionalização de línguas no Brasil: considerações à luz de uma emergente teoria dos direitos linguísticos. *Revista da ABRALIN*, v. 17, n. 2, 30 jun. 2019. Disponível em: <<https://revista.abralin.org/index.php/abralin/article/view/1324>>. Acesso em: 15 jun. 2021.
- \_\_\_\_\_. Prolegômenos para a compreensão dos direitos linguísticos: uma leitura a partir da Constituição da República Federativa do Brasil, In *Sociolinguística e Política Linguística: Olhares Contemporâneos*. São Paulo: Blucher, 2016. Disponível em: <<https://openaccess.blucher.com.br/article-details/prolegomenos-para-a-compreensao-dos-direitos-linguisticos-19936>>. Acesso em: 16 jun. 2021.
- ABREU, P.S.R. Diversidade linguística brasileira, as línguas páno e suas características ergativas. 2008. 336 f. Dissertação (Mestrado em Letras e Linguística) - Universidade Federal de Goiás, Goiânia, 2008. Disponível em: <<http://repositorio.bc.ufg.br/tede/handle/tede/3574>>. Acesso em: 16 jun. 2021.
- ADELAAR, W.F.H. The Endangered Languages Problem: South America. In R.H. Robins & E.M. Uhlenbeck (eds.), *Endangered Languages*. Oxford, New York: Berg. 1991, pp. 45-91. Disponível em: <<https://www.universiteitleiden.nl/en/staffmembers/willem-adelaar#tab-1>>. Acesso em: 16 jun. 2021.
- ALTMAN, M.C.F.S. Trinta anos de Linguística Brasileira e Autoafirmação Profissional". D.E.L.T.A, 1994. Disponível em: <<https://revistas.pucsp.br/delta/article/view/45415>>. Acesso em: 16 jun. 2021.
- ALVES, W.P. Construindo uma Aplicação Web Completa com PHP e MySQL. 1ª. Ed. São Paulo: NOVATEC, 2017.
- AUSTIN, P, K. Data and language documentation. In Gippert, J., Himmelmann, N.P., Mosel, U. *Essentials of Language Documentation*. Berlin: Mouton de Gruyter, 2006.
- \_\_\_\_\_. Los datos y la documentación lingüística In Haviland, J. B; Farfán J. A. F. *Bases de la documentación lingüística*. México, D.F: Instituto Nacional de Lenguas Indígenas, 2007.
- \_\_\_\_\_. Current issues in language documentation. In Peter K. Austin (ed.) *Language Documentation and Description*, vol 7. London: SOAS, 2010. p. 12-33.
- AUSTIN, P, K.; SALLABANK, J. *The Cambridge Handbook of Endangered Languages*. Cambridge: Cambridge University Press, 2011. Disponível em: <<https://www.cambridge.org/core/books/cambridge-handbook-of-endangered-languages/CE451BFE26F3BE24AB78335FD2517EC6>>. Acesso em: 16 jun. 2021.
- AUSTIN, P, K. Language documentation in the 21st century, 2013. Disponível em: <<https://lipp.ub.uni-muenchen.de/lipp/article/view/190>>. Acesso em: 24 set. 2021.
- \_\_\_\_\_. Language documentation 20 years on. In FILIPOVIĆ, L., & PUTZ, M.

Endangered languages and languages in danger: issues of documentation, policy, and language rights, 2015.

AZEVEDO, M. Censos demográficos e "os índios": dificuldades para reconhecer e contar. In: RICARDO, C. A. (Org.) Povos Indígenas no Brasil 1996/2000. São Paulo: Instituto Socioambiental, 2000. p. 79-83. Disponível em: <<https://pib.socioambiental.org/pt/Downloads>>. Acesso em: 16 jun. 2021.

BÁEZ, G. P. et al. Language Documentation and Revitalization in Latin American Contexts. De Gruyter 2016. Disponível: <<http://www.degruyter.com/isbn/9783110428902>>. Acesso em: 29 abr. 2021.

BARRETO, C.T.G., Cardoso, A.M., Coimbra Jr., C.E.A., 2014. Estado nutricional de crianças indígenas Guarani nos estados do Rio de Janeiro e São Paulo, Brasil. Cadernos de Saúde Pública 30, 657–662. Disponível em: <<https://www.scielo.br/j/csp/a/nfr8g6BTYdTSqPcgHHKSbBM/?lang=pt>>. Acesso em: 13 jul. 2021.

BARROS, M. C. D. M. Linguística Missionária: Summer Institute of Linguistics. Tese de Doutorado, Campinas: IFICH / Unicamp, (1993). Disponível em: <<https://www.cpei.ifch.unicamp.br/biblioteca/lingu%C3%ADstica-mission%C3%A1ria-summer-institute-linguistics>>. Acesso em: 16 jun. 2021.

BEVAN, Nigel. Usability issues in website design. 1999. Disponível em: <[https://www.academia.edu/1962551/Usability\\_issues\\_in\\_web\\_site\\_design](https://www.academia.edu/1962551/Usability_issues_in_web_site_design)>. Acesso: 02 nov. 2022.

BLACKMORE, Paul. Intranets: A Guide to Their Design, implementation and Management. London: Taylor & Francis e-Library, 2005.

BOWERN, C. Linguistic Fieldwork: Practical Guide. 2nd ed. Palgrave Macmillian 2015.

BORGES, C.; Erwin, F. Franz Boas: As Limitações do Método Comparativo em Antropologia. Textos e Debates. 2.4 (2012). Disponível em: <<https://revista.ufr.br/textosedebates/article/view/980/0>>. Acesso em: 19 jun. 2021.

BRAGATO, F.F.; NETO, P. B. "Conflitos Territoriais Indígenas No Brasil: Entre Risco e Prevenção / Indigenous Land Conflicts in Brazil: between Risk and Prevention." Revista Direito E Práxis. 8.1, 2017. Disponível em: <<https://www.e-publicacoes.uerj.br/index.php/revistaceaju/article/view/21350/0>>. Acesso em: 12 jul. 2021.

BRASIL. Constituição (1988). Constituição da República Federativa do Brasil. Brasília, DF: Senado, 1988.

BRASIL, Ministério da Educação. As leis e a educação escolar indígena: Programa Parâmetros em ação de Educação Escolar Indígena. Org. Luiz Donisete Benzi Grupioni. Brasília: MEC/SEF, 2001.

BRENZINGER, M. Language Death: Factual and Theoretical Explorations with

Special Reference to East Africa. Berlin: Mouton de Gruyter, 1992.

\_\_\_\_\_. Endangered Languages in Africa. Cologne, Germany: Rüdiger Köper Verlag, 1998.

\_\_\_\_\_. Language Diversity Endangered, Berlin, Boston: De Gruyter Mouton, 2015.

BUSSMANN, H. et al. Routledge Dictionary of Language and Linguistics. Routledge 1998.

CADOGAN, Leon. Las tradiciones religiosas de los Mbya Guarani del Guairá. In Revista de la Sociedade Científica del Paraguay. VII-1. Asunción, 1946.

\_\_\_\_\_. La encarnación y La concepción; La muerte y La resurrección em la poesia sagrada "esotérica" de los Jegukáva-Temondé Porá-Gué (Mbya-Guarani\_ Del Guaíra, Paraguai. Revista do Museu Paulista. v.IV, 1952.

\_\_\_\_\_. Ayvu Rapyta (III). Revista de Antropologia, [S. l.], v. 2, n. 1, p. 37-46, 1954. DOI: 10.11606/2179-0892.ra.1954.110272. Disponível em: <https://www.revistas.usp.br/ra/article/view/110272>. Acesso em: 23 jun. 2021.

\_\_\_\_\_. Ayvu rapyta. Boletim n.227. Antropologia n.5, São Paulo: Universidade de São Paulo, 1959.

CÂMARA JR, Joaquim Mattoso. Introdução às Línguas Indígenas Brasileiras. Rio de Janeiro: Livraria Acadêmica, 1965. Disponível em:

<[http://etnolingustica.wdfiles.com/local--files/biblio%3Acamarajr-1965-introducao/CamaraJr\\_1965\\_IntrodLinguasIndigenasBr.pdf](http://etnolingustica.wdfiles.com/local--files/biblio%3Acamarajr-1965-introducao/CamaraJr_1965_IntrodLinguasIndigenasBr.pdf)>. Acesso em: 16 jun. 2021.

\_\_\_\_\_. História da linguística. Petrópolis: Vozes, 1975.

CAMERON, D., et al. Researching Language: Issues of Power and Method. New York: Routledge, 1992.

CAMPBELL, Lyle. Historical Linguistics: an Introduction. Cambridge, MASS: MIT Press, 2001. pp. 108-162. Disponível em: [https://www.google.com.br/books/edition/Historical\\_Linguistics/EjXrrOJhex8C?hl=pt-BR&gbpv=1&dq=CAMPBELL,+Lyle.+Historical+Linguistics:+an+Introduction&printsec=frontcover](https://www.google.com.br/books/edition/Historical_Linguistics/EjXrrOJhex8C?hl=pt-BR&gbpv=1&dq=CAMPBELL,+Lyle.+Historical+Linguistics:+an+Introduction&printsec=frontcover)>. Acesso em: 16 jun. 2021.

CAMPOS, Carlo Sandro de Oliveira. Morfofonêmica e morfossintaxe do Maxakalí. Tese de doutorado, Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 2009.

CANTONI, G. Stabilizing Indigenous Languages. Flagstaff: North. Ariz. Univ. Cent. Excell. Educ, 1997.

CARDOSO, A.M., Mattos, I.E., Koifman, R.J., 2001. Prevalência de fatores de risco para doenças cardiovasculares na população Guaraní-Mbyá do Estado do Rio de Janeiro. *Cadernos de Saúde Pública* 17, 345–354. doi:10.1590/s0102-311x2001000200009.

CARNEIRO DA CUNHA, Manuela C. Índios No Brasil: História, Direitos E Cidadania. São Paulo: Claro Enigma, 2013.

CASTELLS, M. *Mobile Communication and Society: a Global Perspective: a Project of the Annenberg Research Network on International Communication*. Cambridge, MIT Press. 2004. Disponível em: <<http://ieeexplore.ieee.org/servlet/opac?bknumber=6267235>>. Acesso em: 17 jun. 2021.

CERRÓN-PALOMINO, Rodolfo. *Lingüística Quechua*. Cusco: Centro de Estudios Rurales Andinos "Bartolomé de Las Casas", 1987. Disponível em: <https://www.cambridge.org/core>. UFRJ-SiBI-Sistema de Bibliotecas e Informacao. Acesso em: 28 dez. 2020.

CHAPPEL, H.; LAN, L. *Mandarin and Other Sinitic Languages In Routledge Encyclopedia of the Chinese Language*. Oxford: Routledge, 2017. p. 605-628.

CHAVES, Maria de Betania Garcia. *A política de saúde indígena no município de Angra dos Reis: um estudo de caso*. Rio de Janeiro; s.n; 2006. 156 p. disponível em: <<https://www.arca.fiocruz.br/handle/icict/4641?mode=full>>. Acesso em: 7 jul. 2021.

CHIRIBOGA, Oswaldo Ruiz. *O direito à identidade cultural dos povos indígenas e das minorias nacionais: um olhar a partir do Sistema Interamericano*. Sur. *Revista Internacional de Direitos Humanos* [online]. 2006, v. 3, n. 5. pp. 42-69. Disponível em: <<https://doi.org/10.1590/S1806-64452006000200004>>. Acesso em: 25 set. 2022.

CLASTRES, Pierre. *A Sociedade Contra O Estado: Pesquisas De Antropologia Política*. Rio de Janeiro: F. Alves, 1978.

\_\_\_\_\_. *A Fala Sagrada Mitos E Cantos Sagrados Dos Índios Guarani*. Campinas: Papirus, 1990.

\_\_\_\_\_. *Chronique Des Indiens Guayaki: Ce Que Savent Les Aché, Chasseurs Nomades Du Paraguay*. Paris: Plon, 1991.

COELHO, Rosa J. Plá. *O sistema de proteção das línguas minoritárias na União Europeia. Políticas Culturais em Revista*, 1(6), p. 22-37, 2013. Disponível em: <[www.politicasculturaisemrevista.ufba.br](http://www.politicasculturaisemrevista.ufba.br)>. Acesso em: 15 jun. 2021.

COIMBRA JR., CE.A.; Santos, RV. *Perfil epidemiológico da população indígena no Brasil: considerações gerais*. Centro de Estudo em Saúde do Índio de Rondônia - Cesir, Escola Nacional de Saúde Pública, 2001. Disponível em: <<http://red.bvsalud.org/lis-rede-BVS/resource/17765>>. Acesso em: 05 jul. 2021.

COIMBRA JR., CE.A.; SANTOS, RV; ESCOBAR, AL. (orgs.). *Epidemiologia e saúde dos povos indígenas no Brasil* [online]. Rio de Janeiro: Editora FIOCRUZ; Rio de Janeiro: ABRASCO, 2005. 260 p. ISBN: 85-7541-022-9. Disponível em: SciELO Books <<http://books.scielo.org>>. acesso em: 18 jun. 2021.

COLBY, JOHN. *Practical intranet development*. Apress, 2003.

CONATHAN, Lisa. Archiving and language documentation. In Austin, Peter K. & Julia Sallabank (eds.), *Cambridge Handbook of Endangered Languages*. Cambridge: Cambridge University Press, 2011. p. 235-254.

CORDEIRO, Tássia G.B.Figueiredo. A questão indígena em Maricá: A pesquisa no contexto do Instituto Federal Fluminense. 2016. Disponível em: <<http://www2.fct.unesp.br/docentes/geo/>>. Acesso em: 23 jun. 2021.

CORONEL-MOLINA, S. M.; MCCARTY, T. L.; AUSTIN, P. K.; & SALLABANK, J. Language curriculum design and evaluation for endangered languages, 2011. p. 354-370. Disponível em: <<https://www.cambridge.org/core/books/cambridge-handbook-of-endangered-languages/language-curriculum-design-and-evaluation-for-endangered-languages>> acesso em: 17 jun. 2021.

COSERIU, Eugenio. Premissas históricas da linguística moderna. In \_\_\_\_\_. *Lições de linguística geral*. Rio de Janeiro: Ao Livro Técnico, 1980.

COUTO, H. H. O que vem a ser a ecolinguística, afinal? *Cadernos de Linguagem e Sociedade (UNB)*. 2013. Disponível em: <<https://periodicos.unb.br/index.php/les/article/view/22250>>. Acesso em: 10 abr. 2023.

CREVELS, M. Language endangerment in South America: The clock is ticking. In Campbell, L. and Grondona, V. ed. *The Indigenous Languages of South America*. Berlin, Boston: De Gruyter Mouton, 2012. pp. 167-234. Disponível em: <<https://doi.org/10.1515/9783110258035.167>>. acesso em: 17 jun. 2021.

CRYSTAL, David. *Language Death*. Cambridge, UK: Cambridge University Press, 2000.

CYRANKA, Lucia F. Mendonça. Evolução dos estudos linguísticos. In *Revista Práticas de Linguagem*. v. 4, n. 2, jul./dez. 2014. p. 161-198.

DA SILVA, G.R. Documentação linguística na Amazônia: abordagens e métodos. *Revista Brasileira de Línguas Indígenas*, 2018. Disponível em: <<https://1library.org/subject/linguistica-amazonia>>. Acesso em: 02 ago. 2021.

DALBY, A. 2002. *Language in danger*. London: Penguin. Disponível: <<http://cup.columbia.edu/book/language-in-danger/9780231129008>>. Acesso em: 16 jun. 2021.

DATE, C. J. *Introdução a Sistemas de Bancos de Dados*, Editora Campus, 2004.

DIXON, R. M. W; AIKHENVALD, Alexandra Y. (eds.). The Amazonian languages (Cambridge Language Surveys). Cambridge & New York: Cambridge University Press, 1999. p. 126. Disponível em: <<https://www.jstor.org/stable/4176649>>. Acesso em: 16 jun. 2021.

DIXON, R. The Rise and Fall of Languages. Cambridge: Cambridge Univ. Press, 1997.

DOOLEY, Robert. Vocabulário do Guaraní. Brasília: Summer Instituto f Linguistic, 1982.

DORIAN, N. C. Investigating obsolescence: Studies in language contraction and death. Cambridge University Press. 1989. Disponível em: <<https://www.cambridge.org/core/journals/journal-of-linguistics/article/abs/nancy-n-dorian-ed-investigating-obsolescence-studies-in-language-contraction-and-death-cambridge-cambridge-university-press-1989-pp-xiv-445/33F94848A8A9900BD0B19D200C853FC0>>. Acesso em: 18 jun. 2021.

DOWNING, D. Dictionary of Computer and Internet Terms. Eleventh ed. Barron's Educational Series, 2013.

DRIETRICH, W. O tronco tupi e as suas famílias de línguas. In.: DRIETRICH, W; NOLL, V. O português e o tupi do Brasil. São Paulo: Editora Contexto, 2010.

DUARTE, F. B. Diversidade linguística no Brasil. Revista Caletrosópio. Ouro Preto, MG - UFOP, 2016. Disponível em: <<http://www.letras.ufmg.br/fbonfim/>>. Acesso em: 16 jun. 2021.

DWYER, A.M. Ethics. In Gippert, J., Himmelmann, N.P., Mosel, U. Essentials of Language Documentation. Berlin: Mouton de Gruyter, 2006.

ETHNOLOGUE (2013). Disponível em: <<https://www.ethnologue.com/guides>>. Acesso em: 16 jun. 2021.

ELMASRI, Ramez; NAVATHE, S. B. Sistemas de banco de dados. Revisor técnico Luis Ricardo de Figueiredo. -- São Paulo: Pearson Addison Wesley, 2005.

EPPS, P. et al. Archiving and Language Documentation. Disponível em: <<https://preprints.scielo.org/index.php/scielo/preprint/view/4755/version/5042>>acesso em: 20 mar. 2023.

EVANS, Nicholas. Review of Essentials of Language Documentation. Language documentation and conservation, 2008. p. 340-350.

FAUSTO, Carlos. Os Índios Antes Do Brasil. Rio de Janeiro: Jorge Zahar Editor, 2005.

\_\_\_\_\_. Fragmentos de história e cultura tupinambá: da etnologia como instrumento crítico de conhecimento etno-histórico. In CUNHA, M. C. (org.) História dos Índios no Brasil.

FILL, A; MÜHLHÄUSLER, P. *The Ecolinguistics Reader : Language Ecology and Environment*. London: Continuum International Publishing Group; 2001.

FIORI, José Luís (ORG.): *O Poder Americano*. Petrópolis: Editora Vozes, 2004, 456p.

FISHMAN, J. A. *Language and ethnicity in minority sociolinguistic perspective*. Clevedon: Multilingual Matters. 1988. Disponível em: <https://www.multilingual-matters.com/page/detail/Language-and-Ethnicity-in-Minority-Sociolinguistic-Perspective/?k=9781853590054>. Acesso em: 17 jun. 2021.

\_\_\_\_\_. *Reversing language shift: Theoretical and empirical foundations of assistance to threatened languages*. Clevedon: Multilingual Matters. 1991. Disponível em: <https://eric.ed.gov/?id=ED456659>. Acesso em: 17 jun. 2021.

FRANCHETTO, B.; LEITE, Y. F. *A Concepção dos Linguistas*. Cadernos de Estudos Linguísticos. Márcio Silva (org.). Campinas: Unicamp/Funcamp, 1983.

FRANCHETTO, B. & RICE, K. *Language Documentation in the Americas*. *Language Documentation & Conservation*, 2014, pp. 251-261. Disponível em: <https://scholarspace.manoa.hawaii.edu/handle/10125/24606>. Acesso em: 27 nov. 2021.

FURTADO DA CUNHA, Maria Angélica. *Funcionalismo*. In MARTELOTTA, Mário Eduardo (Org.). *Manual de linguística*. 2. Ed. São Paulo: Contexto, 2011. p. 157-176.

GLASER, K., & HARDEN, J. *A feminist model for ethical decision making*. In E. J. Rave & C. C. Larsen (Eds.), *Ethical decision making in therapy: Feminist perspectives* (pp. 18–37). The Guilford Press, 1995.

GOLLA, V. *The Records of American Indian Linguistics*. in SILVERMAN, Sydel; PAREZO, Nancy J. (editors), *Preserving the Anthropological Record* (2nd ed., New York: 1995). p. 143-157.

GOMES, M. P. *Os índios e o Brasil: Ensaio Sobre um Holocausto e Sobre uma Nova Possibilidade de Convivência*. Petrópolis: Vozes, 1988.

GRENOBLE, L. *Language Policy in the Soviet Union*. Dordrecht: Springer, 2011. Disponível em: <https://www.springer.com/gp/book/9781402012983>. Acesso em: 18 jun. 2021.

GRENOBLE, L; FURBEE, N. *Language Documentation. Practice and values*, 2010. Disponível em: <https://benjamins.com/catalog/z.158>. Acesso em: 18 jan. 2022.

\_\_\_\_\_. *Language Documentation. Practice and values*, 2012. Disponível em: <https://benjamins.com/catalog/z.158>. Acesso em: 10 mar. 2023.

GRENOBLE, L.; SINGERMAN, A. R. *Minority Languages*. In *obo in Linguistics*. 2017. Disponível em: <https://www.oxfordbibliographies.com/view/document/obo-9780199772810/obo-9780199772810-0176.xml>. Acesso em: 15 jun. 2021.

GRENOBLE, L.; WHALEY, L. J. *Endangered Languages: Current Issues and Future Prospects*. Cambridge: Cambridge Univ. Press, 1998. Disponível em: <[https://bridges.monash.edu/articles/journal\\_contribution/Endangered\\_languages\\_current\\_issues\\_and\\_future\\_prospects/5063656](https://bridges.monash.edu/articles/journal_contribution/Endangered_languages_current_issues_and_future_prospects/5063656)>. Acesso em: 16 ago. 2022.

\_\_\_\_\_. *Saving Languages: an Introduction to Language Revitalization*. Cambridge: Cambridge Univ. Press, 2006. Disponível em: <<https://www.cambridge.org/br/academic/subjects/languages-linguistics/sociolinguistics/saving-languages-introduction-language-revitalization?format=HB&isbn=9780521816212>>. Acesso em: 17 jun. 2021.

GRINEVALD, C. *Endangered languages: Language endangerment in South America: a programmatic approach*. 1998. Disponível em: <<https://www.semanticscholar.org/paper/Endangered-languages%3A-Language-endangerment-in-a-Grinevald/3002c5ccf89bad869e219a61383ef39d7d4c6001#citing-papers>>. Acesso em: 17 jun. 2021.

HAGÈGE, C. *Le Souffle de la Langue*. Paris: Odille Jacob, 1992. Disponível em: <[https://www.persee.fr/doc/slave\\_0080-2557\\_1992\\_num\\_64\\_4\\_6095\\_t1\\_0773\\_0000\\_1](https://www.persee.fr/doc/slave_0080-2557_1992_num_64_4_6095_t1_0773_0000_1)> acesso em: 16 ago. 2022.

\_\_\_\_\_. *Halte à la Mort des Langues*. Paris: Editions Odile Jacob, 2000. Disponível em: <<https://www.erudit.org/fr/revues/as/2003-v27-n3-as716/007938ar.pdf>>. Acesso em: 16 ago. 2022.

HAIG, et al. *Documenting Endangered Languages: Achievements and perspectives*. Chapter 4. Berlin: De Gruyter Mouton; 2011.

HALE, Kenneth, et al. "Endangered Languages." *Language*. JSTOR vol. 68, no. 1, 1992, pp. 1–42. Disponível em: <<https://www.jstor.org/stable/416368>>. Acesso em: 16 jun. 2021.

HANSEN, S. et al. *Traditional Knowledge and Intellectual Property : A Handbook on Issues and Options for Traditional Knowledge Holders in Protecting Their Intellectual Property and Maintaining Biological Diversity*. American Association for the Advancement of Science, 2003. Disponível em: <<https://community-wealth.org/sites/clone.community-wealth.org/files/downloads/book-hansen-vanFleet.pdf>>. Acesso em: 25 ago. 2022.

HARMON, D. *In Light of our Differences: How Diversity in Nature and Culture Makes Us Human*. Washington, DC: Smithsonian Inst. Press, 2002. Disponível em: <<https://scholarlypress.si.edu/store/anthropology-archeology/light-our-differences-how-diversity-nature-culture/>>. Acesso em: 16 ago. 2022.

HARRISON, K D. *When Languages Die: The Extinction of the World's Languages and the Erosion of Human Knowledge*. New York: Oxford University Press, 2008. Disponível em: <<https://oxford.universitypressscholarship.com/view/10.1093/acprof:oso/9780195181920.001.0001/acprof-9780195181920>>. Acesso em: 18 jun. 2021.

HAUGEN, E. The ecology of language. The ecology of language. Stanford: Stanford University Press: 325-339, 1972 (também em Fill; Mühlhäusler, pp. 57-66, 2001).

HAZÄEL-MASSIEUX, M.C. Les Créoles: L'Indispensable Survie. Paris: Editions Entente, 1999. Disponível em: <  
<https://bibliotheques.ctguyane.fr/MUSEE/doc/SYRACUSE/101770/les-creoles-l-indispensable-survie-marie-christine-hazael-massieux>>. Acesso em: 16 ago. 2022.

HENKE, R., BEREZ-KROEKER, L. A Brief History of Archiving in Language Documentation, with an Annotated Bibliography. Language Documentation & Conservation, 2016, pp. 411-457. Disponível em:  
 <<https://scholarspace.manoa.hawaii.edu/handle/10125/24714>>. Acesso em: 14 ago. 2021.

HEUSER, C. A. Projeto de Banco de Dados, 5a. ed. Sagra Luzzato, 2004.

HIMMELMANN, N. P. Linguistic Data Types and the Interface between Language Documentation and Description. University of Hawai'i Press, 2012. Disponível em:  
 <<http://hdl.handle.net/10125/4503>>. Acesso em: 19 set. 2021.

\_\_\_\_\_. La documentación lingüística: que es y para que sirve? In Haviland, J. B; Farfán J. A. F. Bases de la documentación lingüística. México, D.F: Instituto Nacional de Lenguas Indígenas, 2007.

\_\_\_\_\_. Language documentation: What is it and what is it good for? In Gippert, J.; Mosel, U. Essentials of Language Documentation. Berlin: Mouton de Gruyter, 2006.

\_\_\_\_\_. Documentary and descriptive linguistics. Berlin: Mouton de Gruyter, 1998. p. 36.

HINTON L; HALE, K. The Green Book of Language Revitalization in Practice. San Diego: Academic, 2001. Disponível em: <  
[https://www.academia.edu/10427280/Leanne\\_Hinton\\_and\\_Ken\\_Hale\\_eds.\\_.\\_The\\_Green\\_Book\\_of\\_language\\_revitalization\\_in\\_practice](https://www.academia.edu/10427280/Leanne_Hinton_and_Ken_Hale_eds._._The_Green_Book_of_language_revitalization_in_practice)>. Acesso em: 16 ago. 2022.

HOBBSAWM, Eric. A Era dos Extremos – O Breve Século XX (1914-1991). São Paulo: Companhia das Letras, 2004.

HOLTON, G. Mediating language documentation. In David Nathan & Peter K. Austin (eds.), Language Documentation and Description, Volume 12: Special Issue on Language Documentation and Archiving, 37–52. London: SOAS. 2014.

IIBA. Um guia para o Corpo de Conhecimento de Análise de Negócios (Guia Babok). Ontario: International Institute of Business Analysis, 2011. Disponível em: <  
<https://www.iiba.org/globalassets/career-resources/business-analysis-resources/iiba-bookstore/files/global-business-analysis-core-standard---portuguese.pdf>>. Acesso em: 05 nov.2022.

INSTITUTO SOCIOAMBIENTAL. Disponível em: <https://www.socioambiental.org/pt->

br/o-isa/programas/povos-indigenas-no-brasil. Acesso em: 18 jun. 2021.

JENSEN, Cheryl. In Dixon, R. e Aïkhenval'd, A. (orgs.) *The Amazonian Languages*. Cambridge, UK: Cambridge University Press, 1999.

JOHNSON, H. Language documentation and archiving, or how to build a better corpus. In Peter K. Austin (ed) *Language Documentation and Description*, vol 2. London: SOAS, 2004. pp. 140-153. Disponível em: <[http://www.e-publishing.org/docs/1/02/ldd02\\_11.pdf](http://www.e-publishing.org/docs/1/02/ldd02_11.pdf)>. Acesso em: 23 set. 2021.

\_\_\_\_\_. Graded access to sensitive materials at the archive of the Indigenous Languages of Latin America, Joint Conference on Digital Libraries, Proceedings., 2003, pp. 176-178. Disponível em: <<https://ieeexplore.ieee.org/document/1204858>>. Acesso em: 25 jan. 2022.

JOSEPH, B. et al. When languages collide: perspectives on language conflict, language competition, and language coexistence, 2003. Disponível em: <<https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.368.3180&rep=rep1&type=pdf>>. Acesso em: 16 ago. 2022.

KAUFMAN, Terrence. The native languages of South America. In Christopher Moseley; R. E. Asher (eds.). *Atlas of the world's languages*, Terrence. London: Routledge. (1994). p. 46-76.

KYMLICKA, W. *Cosmopolitismo, Estado-nación y nacionalismo de las minorías: un análisis crítico de la literatura reciente*. México, Universidad Nacional Autónoma de México. (2001).

KYMLICKA, Will; PATTEN, Alan. *Language Rights and Political Theory*. 2007.

KRAUSS, Michael. The world's languages in crisis. *Language*, vol. 68 no. 1, 1992, p. 4-10. Disponível em: <Project MUSE - Endangered languages: On endangered languages and the safeguarding of diversity (jhu.edu)>. Acesso em: 16 jun. 2021.

KUNG, S. S., & SHERZER, J. The Archive of the Indigenous Languages of Latin America: An Overview. *Oral Tradition*, 2013. Disponível em: <<https://mospace.umsystem.edu/xmlui/handle/10355/65320>>. Acesso em: 20 jan. 2022.

LADEIRA, Maria Inês. *Aldeias livres Guarani do litoral de São Paulo e da periferia da capital*, In *Índios no Estado de São Paulo: resistência e transfiguração*. YANKATU, Comissão Pró-Índio de São Paulo, São Paulo: 1984.

\_\_\_\_\_. *O caminhar sob a luz: Território mbya à beira do oceano*. 1992. Dissertação de Mestrado em Antropologia – Pontifícia Universidade Católica - PUC.

\_\_\_\_\_. *Edição de Publicação*. São Paulo: Editora UNESP, 2007.

\_\_\_\_\_. *Versão Online*. São Paulo: Centro de Trabalho Indigenista –

CTI, 2014.

LAND, Ney; ALMEIDA, Arilza N. População e depopulação em grupos indígenas. *América Indígena* 39 (2). México: Instituto Indigenista Interamericano, 1979. p. 339-369.

LAKATOS, Eva Maria; MARCONI, Marina de Andrade. *Fundamentos de metodologia científica*. 5. ed. São Paulo: Atlas, 2003.

LANZA, Elizabeth; SVENDSEN, BENTE A. Tell me who your friends are and I might be able to tell you what language(s) you speak: Social network analysis, multilingualism, and identity. *International Journal of Bilingualism*, 2007. p. 275-300. Disponível em: <https://journals.sagepub.com/doi/abs/10.1177/13670069070110030201>>. Acesso em: 18 jun. 2021.

LEITE, Yonne. 0 Summer Institute of Linguistics. *Estratégias e Ação no Brasil. Religião e Sociedade* n. 7. Rio de Janeiro, ISER, 1981. pag. 60-64.

LEITE, Yonne; FRANCHETTO, Bruna. “500 anos de línguas indígenas no Brasil”. In Suzana A. M. Cardoso, Jacyra A. Mota, Rosa Virginia Mattos e Silva (orgs.). *Quinhentos Anos de História Linguística do Brasil*. Salvador: Secretaria da Cultura e Turismo do Estado da Bahia, 2006. p. 15-62.

LEMLE, Miriam. Internal classification of the Tupi Guarani linguistic family. In BENDOR-SAMUEL, David (Ed.). *Tupi studies I*. Norman: University of Oklahoma, 1971. (Summer Institute of Linguistics Publications in Linguistics and Related Fields; 29).

LEWIS, M. Paul, Gary F. Simons e Charles D. Fennig (eds) (2014) *Ethnologue: Línguas do Mundo*, Décima sétima edição. Dallas, Texas: SIL International. Disponível em: <http://www.ethnologue.com>. Acesso em: 21 set. 2022.

LINHARES, Maria Yedda. *Prefácio a Cristóbal de Acuña (1641). Novo descobrimento do grande rio Amazonas*. Tradução de Helena Ferreira. Rio de Janeiro: Agir, 1994.

LITAIFF, Aldo. *As divinas palavras: representações étnicas dos Guarani-Mbyá*. (MA thesis, Florianópolis: Universidade Federal de Santa Catarina). 1991. Disponível em: <<https://repositorio.ufsc.br/handle/123456789/75740>>. Acesso em: 23 jun. 2021.

LIZARRALDE, Manuel. Biodiversity and loss of Indigenous languages and knowledge in South America. In Maffi, Luisa (ed.) *On biocultural diversity: Linking language, knowledge, and the environment*. Washington: Smithsonian Institution Press. p. 265–281.

LÜPKE, F. Data collection methods for field-based language documentation. In Peter K. Austin (ed.) *Language Documentation and Description*, vol 6. London: SOAS, (2010), pp. 53-100. Disponível em: <<http://www.elpublishing.org/PID/071>>. Acesso em: 20 set. 2021.

LYONS, John. *Language and Linguistics: An Introduction*. Cambridge: Cambridge University Press, 1981.

MACHADO, F.N.R. *Projeto de Banco de Dados: uma visão prática*. 15a Ed. São Paulo: Érica, 2008.

\_\_\_\_\_. *Projeto de Banco de Dados: uma visão prática*. 8a Ed. São Paulo: Érica, 2004.

MACKEY, W. F. The ecology of language shift. in FILL, A. & MÜHLHÄUSLER, P. *The ecolinguistics reader*. London: Continuum, [1980] 2001. p. 67–74.

MAFFI, L. *On Biocultural Diversity: Linking Language, Knowledge, and the Environment*, 2001. Disponível em: <[https://www.sil.org/system/files/reapdata/10/87/34/108734828518823365457562969695287058993/SILEBR\\_2004\\_007.pdf](https://www.sil.org/system/files/reapdata/10/87/34/108734828518823365457562969695287058993/SILEBR_2004_007.pdf)>. Acesso em: 16 ago. 2022.

\_\_\_\_\_. *The business of language endangerment: Saving languages or helping people keep them alive*. 2003. Disponível em: <<https://benjamins.com/catalog/wlp.1.07maf>>. Acesso em: 18 jun. 2021.

MAIA, Marcus; FRANCHETTO, Bruna; LEMLE, Miriam; VIEIRA, Márcia D. *Línguas Indígenas e Gramática Universal*. 2019.

MARGETTS, A. & MARGETTS, A. In THIEBERGER, N. *The Oxford handbook of linguistic fieldwork*, 2014. Disponível em: <<https://www.vlebooks.com/vleweb/product/openreader?id=none&isbn=9780191632815>>. Acesso em: 22 set. 2021.

MARGETTS, A. Data processing and its impact on linguistic analysis. *Language Documentation & Conservation*, 2009, pp. 87-99. Disponível em: <<https://core.ac.uk/download/pdf/5098477.pdf>>. Acesso em: 21 set. 2021.

MARTINET, A. *Elementos de lingüística geral*. Trad. Jorge Morais Barbosa. 5. Ed. Lisboa: Sá da Costa, 1973.

MARTIUS, Karl Friedrich Phillipp von. *Beiträge zur Ethnographie und prachenkunde Amerika's zumal Brasiliens*, I. Zur Ethnographie, II. Zur Sprachenkunde. Leipzig: Friedrich Fleischer, 1867.

MAURAS, J; MORRIS, M.A. *Languages in a Globalizing World*. Cambridge: Cambridge Univ. Press, 2003. Disponível em: <<http://catdir.loc.gov/catdir/samples/cam033/2002067728.pdf>>. Acesso em: 16 ago. 2022.

MELATTI, Julio C. *População indígena*. Brasília: Departamento de Antropologia, Universidade de Brasília. 2004. Disponível em: <<http://dan.unb.br/dan-producaocientifica/serieantropologia>>. Acesso em: 18 jun. 2021.

\_\_\_\_\_. Índios do Brasil. São Paulo: EDUSP, 2007. Disponível em: <<https://www.edusp.com.br/livros/indios-do-brasil/>>. Acesso em: 17 jun. 2021.

MELIÁ, Bartolomeu. A Experiência Religiosa dos Guarani. In O Rosto índio de Deus. São Paulo: Editora Vozes. 1989.

\_\_\_\_\_. "Indigenous Cultures and Evangelization: Challenges for a Liberating Mission." *International Review of Mission*. 81.324, 1992. p. 557-567.

MELLO, A. A. S.; KNEIP, A. Babel indígena. *Revista de História*, v.8., Rio de Janeiro, 2013. p. 21-23

MÉTRAUX, Alfred. A religião dos Tupinambás. São Paulo: Ed. Nacional: Edusp, 1979.

MONTERO, F., Lozano, M., Gonzáles, P. and Ramos, I. "A First Approach To Design Web Sites By Using Patterns", *Proceedings of VikingPLoP Conference*, 2002.

Disponível em: <

[https://www.researchgate.net/publication/266871459\\_A\\_First\\_Approach\\_To\\_Design\\_Web\\_Sites\\_By\\_Using\\_Patterns](https://www.researchgate.net/publication/266871459_A_First_Approach_To_Design_Web_Sites_By_Using_Patterns)>. Acesso em: 03 nov. 2022.

MONTEIRO, J. M. A dança dos números - A população indígena do Brasil desde 1500. *Tempo e Presença*. Rio de Janeiro e São Paulo. 1994.

MOORE, D. Endangered Languages of Lowland Tropical South América. In Brenzinger, Matthias. *Language Diversity Endangered*. Berlin: Mouton de Gruyter, 2007. p. 29-58. Disponível em:

<<https://www.degruyter.com/document/doi/10.1515/9783110197129.29/html>>.

Acesso em: 17 jun. 2021.

\_\_\_\_\_. As línguas indígenas no Brasil hoje. In MELLO, H.; ALTHENHOFEN, C. e RASO, T. (Orgs.) *Os Contatos Linguísticos No Brasil*. 1ª ed. Belo Horizonte: Editora UFMG, 2011. p. 217-240.

MOORE, D. & GALUCIO, A. Perspectives for the documentation of indigenous languages in Brazil In G. Pérez Báez, C. Rogers & J. Rosés Labrada (Ed.), *Language Documentation and Revitalization in Latin American Contexts*. Berlin, Boston: De Gruyter Mouton, 2016, pp. 29-58. Disponível em:

<<https://doi.org/10.1515/9783110428902-002>>. Acesso em: 27 nov. 2021.

MORENO FERNÁNDEZ, F. La diversidad lingüística de Hispanoamérica: implicaciones sociales y políticas. *Real Instituto Elcano de Estudios Internacionales y Estratégicos*. 2006. Disponível em:

<<http://dialnet.unirioja.es/servlet/oaiart?codigo=1984397>>. Acesso em: 18 jun. 2021.

MORI, A. C. Teses e Dissertações sobre Línguas Indígenas apresentadas no Instituto de Estudos da Linguagem (IEL – UNICAMP): período 1977-2008. *Liames*, Campinas, v. 7, nº 1, 2007. p. 109-123.

MOSEL, U. Putting oral narratives into writing – experiences from a language

documentation project in Bouganville, Papua New Guinea. In Comrie B. & Golluscio Lucía. Language Contact and Documentation. Berlin Germany: De Gruyter: Mouton; 2014.

\_\_\_\_\_. Fieldwork and community language work. In Gippert, J., Himmelmann, N.P., Mosel, U. Essentials of Language Documentation. Berlin: Mouton de Gruyter, 2006.

MOSELEY, C. The UNESCO Atlas of the World's Languages in Danger: Context and Process, 2012. Disponível em: <<http://www.dspace.cam.ac.uk/handle/1810/243434>>. Acesso em: 17 jun. 2021.

MUFWENE, S.S. The Ecology of Language Evolution. Cambridge: Cambridge Univ. Press, 2001. Disponível em: <[https://assets.cambridge.org/97805217/91380/frontmatter/9780521791380\\_frontmatter.pdf](https://assets.cambridge.org/97805217/91380/frontmatter/9780521791380_frontmatter.pdf)>. Acesso em: 16 ago. 2022.

MUFWENE, S.S. Language Birth and Death. Annual Review of Anthropology, 2004 33:1, 201-222. Disponível em: <<https://www.annualreviews.org/doi/abs/10.1146/annurev.anthro.33.070203.143852>>. Acesso em: 16 out. 2021.

MÜHLHÄUSLER, P. Linguistic Ecology: Language Change and Linguistic Imperialism in the Pacific Region. London: Routledge, 1996. Disponível em: <<https://www.jbe-platform.com/content/journals/10.1075/sl.22.3.11lic>>. Acesso em: 16 ago. 2022.

MULLINS, C. Database administration: The complete guide to practices and procedures. Upper Saddle River, NJ: Addison-Wesley, 2002.

NATHAN, D. Sound and unsound practices in documentary linguistics: towards an epistemology for audio, vol 7, 1-17. London: SOAS. 2010.

NETTLE, D.; ROMAINE, S. Vanishing voices: The extinction of the world's languages. New York: Oxford University Press. 2000.

NEWMAN, P. Fieldwork and field methods in linguistics. Language Documentation & Conservation, 2007. p. 283-289.

NIEDERAUER, J. Desenvolvendo websites com PHP. 3ª Ed. São Paulo: NOVATEC, 2017.

NIMUENDAJÚ, Curt Unkel. As Lendas da Criação e Destruição do Mundo, como fundamentos da religião dos Apapocúva - guarani. São Paulo: Editora HUCITEC-USP. 1987.

ORLANDI, Eni Puccinelli. O que é linguística? Ed. Brasiliense. 15. ed. São Paulo: Brasiliense, 2009.

OITICICA, J. Do Método de Estudos das Línguas Sul-Americanas. Boletim SEKI 290 D.E.L.T.A., Vol. 15, Nº ESPECIAL do Museu Nacional, 9. Rio de Janeiro: Museu

Nacional, 1933. p. 41-81.

OMONIYI, Tope; WHITE, Goodith. *The Sociolinguistics of Identity*. London: Continuum, 2008.

ORTIZ, R. Ivarra; MACHADO, A. Martins. Tembiasakue Rapê: a longa estrada Guarani na história e na memória- reconstruindo o passado, ressignificando o presente e trilhando o futuro. *Tellus* ano 18, n. 37, set./dez. 2018. Disponível em: <<https://www.tellus.ucdb.br/tellus/article/view/525>>. Acesso em: 23 jun. 2021.

OWENS, Jonathan. *Arabic as a Minority Language*, Berlin, Boston: De Gruyter Mouton, 2013. Disponível em: <<https://doi.org/10.1515/9783110805451>>. Acesso em: 10 out. 2020.

PARKER, Gary John. 1969. *Ayacucho Quechua Grammar and Dictionary*. The Hague: Mouton. 1969.

PÉREZ BÁEZ, Gabriela; ROGERS, Chris; LABRADA, Jorge Emilio Rosés. *Language Documentation and Revitalization in Latin American Contexts*. Berlin, Boston: De Gruyter Mouton, 2016. Disponível em: <<https://doi.org/10.1515/9783110428902>>. Acesso em: 16 jun. 2021.

PETTER, Margarida. Linguagem, língua, lingüística. In *Introdução à linguística*. FIORIN, José Luiz (org.). 6ª ed. São Paulo: Contexto, 2010.

PISSOLATO, Elizabeth. *A Duração da Pessoa: Mobilidade, parentesco e xamanismo Mbya (Guarani)*. São Paulo, SP: Editora UNESP, 2007.

PHILLIPSON, R. *English Only? Challenging Language Policy*. London: Routledge, 2003. Disponível em: <<https://www.taylorfrancis.com/books/mono/10.4324/9780203696989/english-europe-robert-phillipson>>. Acesso em: 16 ago. 2022.

POMPILHO, S. *Análise Essencial. Guia Prático de Análise de Sistemas*. São Paulo: IBPIPRESS, 1995.

REYHNER, J, Cantoni G, St. Clair RN, Yazzie EP. Revitalizing indigenous languages. *Flagstaff: North. Ariz. Univ. Cent. Excell. Educ*, 1999. Disponível em: <<https://jan.ucc.nau.edu/~jar/RIL.pdf>>. Acesso em: 17 ago. 2022.

RIBEIRO, Darcy. *O povo brasileiro. A formação e o sentido do Brasil*. São Paulo: Companhia das Letras, 1997 (1ª. ed.: 1995).

RIBEIRO, E. R. *A Grammar of Karaja*, Tese de Doutorado, 2012. Disponível em: <<https://www.worldcat.org/title/grammar-of-karaja/oclc/840636974>>. Acesso em: 18 jan. 2022.

\_\_\_\_\_. *Análise e descrição da língua Ofayé*, Dissertação de Mestrado, 2006. Disponível em: <<http://www.etnolingüística.org/resumo:2>>. Acesso em: 18 jan. 2022.

RICARDO, Carlos Alberto (ed.). Povos Indígenas no Brasil, 1996-2000. São Paulo: Instituto Socioambiental. 832 pp. (mapas, ilustrações, gráficos, tabelas).

RICE, K. Ethical Issues in Linguistic Fieldwork: an Overview. *Journal of Academic Ethics*. 4 (2006): p. 123-155.

RICENTO, Thomas; PELED, Yael; IVES, Peter. *Language policy and Political Theory: building bridges, assessing breaches*. Switzerland, 2015.

RIBEIRO, M. et al. Projeto de Sistemas Web. Cuiabá: UFMT-MEC, 2015. Disponível em: < <http://proedu.rnp.br/bitstream/handle/123456789/1536/87.Projeto%20Sistemas%20Web%20-%20INFORM%C3%81TICA%20-%20IFRO.pdf?sequence=1&isAllowed=y>>. Acesso em: 01 nov. 2022.

ROBINS, R. H.; UHLENBECK, E. M. *Endangered Languages*. Providence, RI: Berg, 1991.

RODRIGUES, A. D. Os Estudos de Linguística Indígena no Brasil. *Revista de Antropologia*, separata do vol. XI, nº ½. São Paulo, 1963.

\_\_\_\_\_. Contribuições das línguas brasileiras para a fonética e a fonologia. In SOLÁ, D. F. (Org.) *Language in the Americas*. Ithaca: Cornell University, 1984.

\_\_\_\_\_. Relações internas na família linguística tupi-guarani. *Revista de Antropologia*, v. 27/28, 1985. p. 33-53.

\_\_\_\_\_. Línguas Indígenas: 500 anos de Descobertas e Perdas. *D.E.L.T.A.*, 83-103, 1993.

\_\_\_\_\_. *Línguas brasileiras – para o conhecimento das línguas indígenas*. São Paulo: Edições Loyola, 1994. (1ª. ed.: 1986).

\_\_\_\_\_. Macro-Jê. In DIXON, RMW; AIKHENVALD, Alexandra Y. (Org.). *The amazonian languages*. Cambridge: Cambridge University Press, 1999.

\_\_\_\_\_. Hipóteses sobre as migrações dos três subconjuntos meridionais da família Tupi-Guarani. In CONGRESSO NACIONAL DA ABRALIN, 2, 2000. Atas. Florianópolis: UFSC, 2000. CD-ROM.

\_\_\_\_\_. Sobre as línguas indígenas e sua pesquisa no Brasil. In *Ciência e Cultura*, v.57, n.2, São Paulo, 2005.

\_\_\_\_\_. 40 anos de linguística, cursos universitários e línguas indígenas no Brasil: vivências e memórias pessoais [40 years of linguistics, university courses, and indigenous languages in Brazil: experiences and personal memories]. In Dermeval da Hora, Eliane Ferraz Alves & Lucienne C. Espindola (eds.), *ABRALIN: 40 anos em cena*, 17–34. João Pessoa: Editora Universitária, 2009.

\_\_\_\_\_. *A originalidade das línguas indígenas brasileiras*. Brasília, DF:

Laboratório de Línguas Indígenas, 2013. Disponível em: <<http://www.laliunb.com.br>>. Acesso em: 18 jun. 2021.

RUSSELL, Jesse; COHN, Ronald. Linguistic rights. Scotland: Bookvika publishing, 2012.

RUSSO, K.; Valeriano, I. "Vocês não sabem sentir a palavra!": desafios para a escola indígena do povo Guarani Mbya no Rio de Janeiro. Impulso, Piracicaba • 26(66), 51-66, maio-ago. 2016. Disponível em: <<https://www.metodista.br/revistas/revistas-unimep/index.php/impulso/article/view/2699>>. Acesso em: 10 jul. 2021.

SADDY, André. Códigos de conduta e boas práticas. *Revista de informação legislativa*: RIL, v. 54, n. 215, p. 27-57, jul./set. 2017. Disponível em: <[https://www12.senado.leg.br/ril/edicoes/54/215/ril\\_v54\\_n215\\_p27](https://www12.senado.leg.br/ril/edicoes/54/215/ril_v54_n215_p27)>. Acesso em: 22 nov. 2021.

SALLABANK, Julia. Attitudes to Endangered Languages: Identities and policies. Cambridge University Press, 2013. Disponível em: <<https://www.cambridge.org/core/books/attitudes-to-endangered-languages/E307433DF7990584EDADE97C74D698B9>>. Acesso em: 17 jun. 2021.

SANTANA, Áurea Cavalcante. Descrição de línguas indígenas: histórico, desafios e perspectivas. Polifonia, Cuiabá-MT, v. 25, 38.1, p. 73-89, maio-agosto. 2018.

SAPIR, Edward. Culture, Language and Personality. University of California Press, 1970.

SAUSSURE, F. de. Curso de Lingüística Geral. Tradução Antônio Chelini, José Paulo Paes, Isidoro Blikstein. 25.ed. São Paulo: Cultrix, 1999.

SAWALA, M.R. Dicionário de Informática e Internet. São Paulo: Nobel, 1999.

SCHADEN, Egon. Aspectos fundamentais da cultura Guarani. 3.ed. São Paulo: E.P.V./Edusp, 1974.

SCHADEN, Egon. A Mitologia Heróica De Tribos Indígenas Do Brasil: Ensaio Etnossociológico. Rio de Janeiro: Ministerio da educação e cultura, 1989.

\_\_\_\_\_. Aspectos Fundamentales De La Cultura Guaraní. Asunción: Universidad Católica, 1998.

SCHULTZE-BERNDT, E. "60. Language Documentation". Volume 3, edited by Tibor Kiss and Artemis Alexiadou, Berlin, München, Boston: De Gruyter Mouton, 2015, pp. 2063-2095. Disponível em: <<https://doi.org/10.1515/9783110363685-020>>. Acesso em: 20 nov. 2021.

SEKI, L. Perspectivas para os estudos linguísticos no Brasil. Boletim da ABRALIN. Campinas: IEL/Unicamp, 1991.

\_\_\_\_\_. A linguística indígena no Brasil. Revista D.E.L.T.A., vol. 15, n.

ESPECIAL, p. 257-290, 1999.

\_\_\_\_\_. Línguas Indígenas do Brasil no limiar do século XXI. Revista Impulso, Piracicaba – SP. Vol. 12, n. 27, 2000 p. 157-170.

SEIFART, F. On the representativeness of language documentations. In Peter K. Austin (ed.) Language Documentation and Description, vol 5. London: SOAS, 2008, pp. 60-76. Disponível em: <[http://www.elpublishing.org/docs/1/05/lld05\\_04.pdf](http://www.elpublishing.org/docs/1/05/lld05_04.pdf)>. Acesso em: 19 set. 2021.

SEMENOV, A. Information and communication technologies in schools: a handbook for teachers, or how ICT can create new, open learning environments, 2005. Disponível em: <<https://unesdoc.unesco.org/ark:/48223/pf0000139028>>. Acesso em: 22 mar. 2022.

SENGUPTA, D., & SAHA, G. Identification of the major language families of India and evaluation of their mutual influence. Current Science, 110(4), (2016): 667–681. Disponível em: <http://www.jstor.org/stable/24907928>. Acesso em: 21 set. 2022.

SERAFIM, M.G. Hábitos Alimentares e Nível de Hemoglobina em Crianças Indígenas Guarani, Menores de 5 Anos, dos Estados de São Paulo e do Rio de Janeiro. Dissertação de Mestrado, São Paulo: Escola Paulista de Medicina. 1997.

SILBERSCHATZ, A.; KORTH, H.F.; SUDARSHAN, S.: Sistema de Banco de Dados, tradução da 5a. ed. [por] Daniel Vieira, Rio de Janeiro, Elsevier, 2006.

SINGERMAN, R. The Morphosyntax of Tuparí, a Tupían Language of the Brazilian Amazon, 2018. Disponível em: <<https://adamsingerman.wordpress.com/>>. Acesso em: 18 jan. 2022.

SKUTNABB-KANGAS, T. Linguistic Genocide in Education—or World-Wide Diversity and Human Rights? Mahwah, NJ: Lawrence Erlbaum, 2000. Disponível em: <<https://www.taylorfrancis.com/books/mono/10.4324/9781410605191/linguistic-genocide-education-worldwide-diversity-human-rights-tove-skutnabb-kangas>>. Acesso em: 17 ago. 2022.

SOARES, M.F., LEITE, Y. Vowel shift in the Tupi-Guarani language family: a typological approach. In KEY, M. R. (Ed.). Language change in South American indian languages. Philadelphia: University of Pennsylvania Press, 1991. p. 36-53.

SOARES, M.F. Línguas indígenas: considerações sobre o Acordo Funai-Max Planck, 2007.

SOMMERVILLE, I. Ingeniería de Software. PEARSON Educación: México, 2011.

SOTO, V.V. Documenting different genres of oral narrative in Cora (Uto-Aztecan). In GRENOBLE, L & FURBEE-LOSEE, L. Language Documentation: Practice and Values. Pbk. ed. Amsterdam The Netherlands: John Benjamins Pub; 2012.

STORTO, Luciana. Línguas Indígenas: Tradição, Universais e Diversidade, 2019.

SULLIVAN, R. Indigenous Cultural and Intellectual Property Rights: a Digital library context. *D-Lib Magazine*. May, Vol. 8 No 5, 2002. Disponível em: <<http://www.dlib.org/dlib/may02/sullivan/05sullivan.html>> Acesso em: 22 ago. 2022.

SUTHERLAND, W. J. Parallel extinction risk and global distribution of languages and species, 2003. *Nature* 423: 276–79.

SWAAN, A. *Words of the World: The Global Language System*. Cambridge, UK: Polity Press, 2001. Disponível em: <<https://www.wiley.com/en-us/Words+of+the+World:+The+Global+Language+System-p-9780745627472>>. Acesso em: 17 ago. 2022.

TAKAI, O.K. *Introdução a Banco de Dados*. São Paulo: IME-USP, 2005.

TEOREY, Tobey J. *Projeto e modelagem de banco de dados / Tobey J. Teorey [et al.]; tradução Daniel Vieira. – 2. ed. – Rio de Janeiro: Elsevier, 2014.*

THIEBERGER, N. Report on the AIATSIS visiting research fellowship, Aboriginal Studies Electronic Data Archive: A report to AIATSIS Council on the conclusion of the Visiting Research Fellowship, 1994. Disponível em: <<http://trove.nla.gov.au/work/33785959?q&versionId=41559386>>. Acesso em: 22 ago. 2022.

THIEBERGER, N, & MUSGRAVE, S. *Documentary linguistics and ethical issues*. Hans Rausing Endangered Languages, 2007. Disponível em: <<http://arrow.monash.edu.au/hdl/1959.1/217006>>. Acesso em: 10 dez. 2021.

THOMASON, Sarah G. *Language contact*. Edinburgh: Edinburgh University Press, 2001.

TRASK, R.L.; Stockwell, Peter. *Language and Linguistics: The Key Concepts* Abingdon, Oxon: Routledge, 2007.

TRILSBEEK, P.; WITENBURG, P. Los acervos lingüísticos digitales y sus desafíos In Haviland, J. B; Farfán J. A. F. *Bases de la documentación lingüística*. México, D.F: Instituto Nacional de Lenguas Indígenas, 2007.

\_\_\_\_\_. Archiving challenges In Gippert, J., Himmelmann, N.P., Mosel, U. *Essentials of Language Documentation*. Berlin: Mouton de Gruyter, 2006.

TRILSBEEK, P; WOOD, N. *MPI for Psycholinguistics*, 2008. Disponível em: <<https://www.mpi.nl/corpus/a4guides/a4-guide-TMPGEnc.pdf>>. Acesso em: 25 ago. 2021.

WARDHAUGH, R. *Languages in competition: dominance, diversity, and decline*. Oxford [England]: B. Blackwell in association with A. Deutsch, 1987.

WHALEN, D. How the study of endangered languages will revolutionize linguistics. In

Piet van Sterkenburg (ed.), *Linguistics Today – Facing a greater challenge*. Amsterdam: John Benjamins, 2004. p. 321-342.

WIDLAK, T. The Archive Strikes Back: Effects of Online Digital Language Archiving on Research Relations and Property Rights. In Ameka, F. et al. *Oral Literature in the Digital Age: Archiving Orality and Connecting with Communities*. Cambridge: Open Book Publishers, 2015. Disponível em: <<http://books.openedition.org/obp/1264>>. Acesso em: 26 fev. 2022.

WILKINS, D. P. Linguistic research under Aboriginal control: a personal account of fieldwork in Central Australia. *Australian Journal of Linguistics*, 1992. p. 171-200.

WITTENBURG, P., Mosel, U., & Dwyer, A. Methods of Language Documentation. In the DOBES project. LREC, 2002.

WITTENBURG, P. Code of Conduct DOBES, 2005. Disponível em: <[https://dobes.mpi.nl/ethical\\_legal\\_aspects/DOBES-coc-v2.pdf](https://dobes.mpi.nl/ethical_legal_aspects/DOBES-coc-v2.pdf)>. Acesso em: 15 nov. 2021.

WOODBURY, A. C. Defining documentary linguistics. In Peter K. Austin (ed.) *Language Documentation and Description*, vol 1. London: SOAS, 2003. p. 35-51.

WOODBURY, A. C. Language documentation. In Peter K. Austin & Julia Sallabank (eds.), *The Cambridge Handbook of Endangered Languages*, 159-186. Cambridge: Cambridge University Press, 2011.

WHORF, Benjamin Lee; CARROLL, John B. *Language, Thought and reality: selected writings*. [Cambridge, Mass.]: Technology Press of Massachusetts Institute of Technology; New York: John Wiley & Sons, 1956.

WURM, S.A. *Language Atlas of China – A Project*. ASIEN, v. 14, 1985. p. 80-87.

ZEPHIRO, K.A.; MARTINS, N.D.J., 2015. Educação Escolar Indígena Diferenciada e Intercultural entre os Guarani Mbyá do Rio de Janeiro: o legítimo e o real. *REAL. Revista Periferia*, v. 7, n. 1, jan./jun. 2015. Disponível em: <<https://www.e-publicacoes.uerj.br/index.php/periferia/article/view/21968>>. Acesso em: 13 jul. 2021.

## ANEXOS

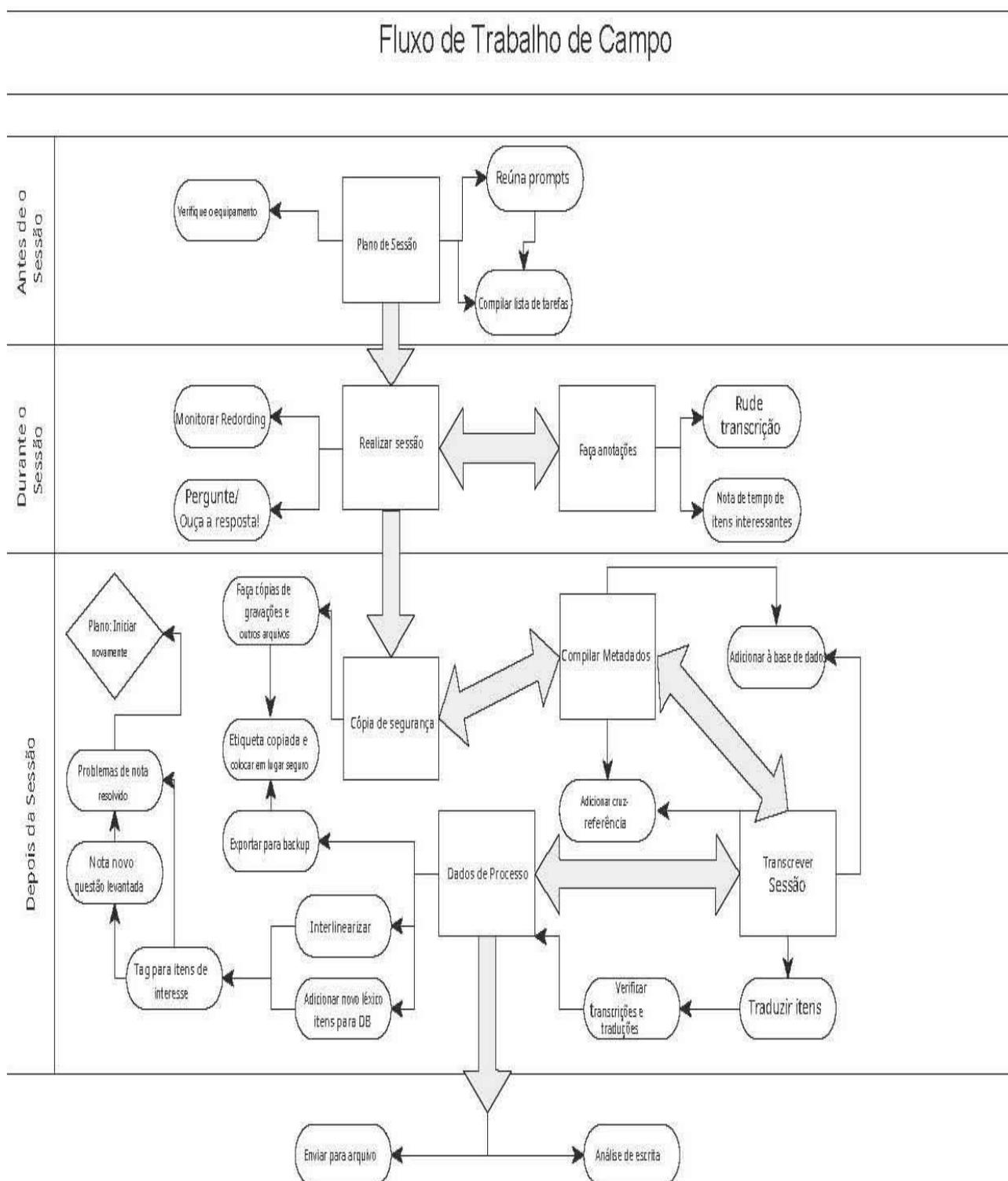
ANEXO A - Kit básico de equipamento de campo<sup>173</sup>

QTD	EQUIPAMENTO	ACESSÓRIOS EXTRAS
1	Gravador digital portátil tipo mp3	-
1	Gravador de áudio de estado sólido (SSR) <sup>174</sup>	4-6 baterias; 1-2 carregadores de bateria
2	Microfones (estéreo, de lapela ou com pedestal)	Divisor de microfone (permite uso simultâneo de 2 microfones)
2	Fones de ouvido (alta fidelidade)	Divisor de fone de ouvido (permite uso simultâneo de 2 fones)
1	Câmera de vídeo profissional	4-6 baterias; 1-2 carregadores de bateria; tripé, punho e / ou estabilizador de câmera
1	Câmera de vídeo auxiliar	Para backup e documentação de eventos com duas câmeras
2	Computadores portáteis (o mais robusto possível), um para uso constante e outro como auxiliar em caso de pane do principal	2 baterias; cabo de alimentação 12 volts; pendrives e/ou hd externo; cabos (XLR para microfone do gravador / câmera; USB / Fireware para transferência / captura no computador)
1	Dispositivo GPS	-
1	Câmera fotográfica digital	-
1	Suporte para laptop ou teclado externo	-
1	Painel solar	-
1	Bateria de carro / caminhão, fusíveis e cabo	-
1	Multímetro, ferro de solda e ferramentas básicas de eletricista	-
-	Sacos impermeáveis (pelo menos um para cada equipamento)	-
-	Plástico bolha (para enchimento de bolsas)	-

<sup>173</sup> Quadro retirado e adaptado de Bowern (2015, p. 18-36) e Margetts & Margetts (2012, p. 56-57)

<sup>174</sup> Cf. Gravador de áudio de estado sólido (SSR). Disponível em: <<https://definirtec.com/gravador-de-estado-solido-ssr/>>. Acesso em: 13 set. 2021.

## ANEXO B - Fluxograma etapas da documentação linguística no trabalho de campo<sup>175</sup>



<sup>175</sup> Figura retirada e adaptado de Bowern (2015, p. 53)

## ANEXO C - Guia DOBES para gravação de áudio<sup>176</sup>

### Fazendo gravações de áudio

Comentários e dicas para gravar áudio

#### Requisitos:

- Aplicativo de gravação de áudio. (MD rec., Flashcardrec. Filmadora)
- Manuais para os dispositivos.
- Mídia vazia suficiente.
- Baterias sobressalentes suficientes ou uma fonte de alimentação separada para o gravador
- Microfone (s) (dependendo da situação: pequenos tripés, pára-brisas, etc.)
- Fones de ouvido pequenos para verificações de som.
- Para digitalização: uma porta USB com um dispositivo de áudio ou uma placa de som para processamento
- Um programa de digitalização apropriado: 'Adobe Audition', 'Audacity' ou similar.

#### 1. Geral

- Familiarize-se com seu equipamento de gravação e digitalização.
- Consulte os manuais se algo não estiver claro e também para garantir que você está operando os dispositivos corretamente
- Verifique se você possui uma fonte de alimentação totalmente carregada e estável ou baterias novas.
- Antes de começar com a gravação "real", faça uma gravação de teste para verificar se tudo está configurado e conectado corretamente.
- Fale e grave o texto do ID da etiqueta e rotule a mídia (cassete ou fita). Isso é necessário para identificação posterior.
- Faça uma verificação de som na gravação do teste.

#### 2). Gravando áudio

Se você precisar gravar ao ar livre, tente encontrar um lugar calmo e protegido ao vento. Se necessário, use protetor de vento para o (s) microfone (s).

Não exponha o equipamento à luz solar por muito tempo, pois os dispositivos podem esquentar para que as especificações fornecidas não sejam mais válidas. Tente encontrar um lugar sombreado.

Coloque o microfone o mais próximo possível, mas não muito próximo do ponto de interesse.

Esteja ciente das características do microfone. Verifique se o microfone não "ouve" o chão ou o rosto na direção errada.

Não use cabos de extensão longa. Mantenha o cabo do microfone o mais curto possível. Devido à construção física, os microfones mais "assimétricos" são sensíveis ao zumbido e outros ruídos. Os microfones "simétricos" não sofrem com isso, mas nem sempre são compatíveis com o equipamento de gravação padrão.

Mantenha as superfícies do microfone limpas. Uma mosca ou uma abelha pode perturbar completamente a gravação ao voar ao redor do microfone, porque cheira bem.

Defina o nível de gravação do dispositivo para que o sinal não seja "registrado demais" e também não seja muito fraco. Especialmente para gravadores digitais, gravar muito alto dá um som distorcido feio. -3db é um bom nível de pico.

As gravações em mídia digital devem ser uma taxa de amostragem de 44,1 Khz ou 48 Khz e uma resolução de pelo menos 16 bits. (também chamado de gravação linear de PCM)

Não faça gravações no formato MP3 ou ATRAC (MiniDisk) para arquivamento posterior, a menos que não haja realmente outra opção.

Nunca grave no modo LP (Long Play) com dispositivos digitais. Isso produzirá arquivos de áudio não lineares. (veja a folha separada de Técnicas de compressão de áudio.)

<sup>176</sup> Figura retirada e adaptada de The Language Archive (MPI for Psycholinguistics). Disponível em: <<https://www.mpi.nl/corpus>>. Acesso em: 16 set. 2021.

## ANEXO D - Guia DOBES para gravação de vídeo – parte 1<sup>177</sup>

### Fazendo gravações de vídeo

Comentários, dicas e dicas para gravar vídeo

#### Requisitos:

- Câmera (câmera de vídeo digital)
- Manuais para os dispositivos.
- Mídia vazia suficiente.
- Baterias sobressalentes suficientes ou uma fonte de alimentação separada para o gravador.
- Tripé extra-externos de microfone, pára-brisas, cabos de extensão, lentes grande angulares, fita de limpeza para a câmera de vídeo, etc.)
- Faróis pequenos para verificações de som.
- Para digitalização: uma porta IEEE ('fio de fogo') no PC / laptop
- Um programa de digitalização apropriado: "Adobe Premiere" ou "Adobe Elements".

#### 1. Geral

- Familiarize-se com seu equipamento de gravação e digitalização. É muito importante que se treine para fazer gravações de alta qualidade.
- Consulte o (s) manual (s) se algo não estiver claro e também para garantir que você esteja operando os dispositivos corretamente.
- Mantenha seu equipamento em local seco quando não estiver em uso. A umidade pode tornar a gravação impossível.
- Verifique se você possui uma fonte de alimentação estável totalmente carregada ou baterias novas.
- Faça uma gravação curta de teste. Escreva o ID da etiqueta e outras informações relevantes em uma folha de papel branco e grave-a, leia também o ID da etiqueta em voz alta para reconhecimento de áudio. Isso é útil para identificação posterior. Rotule a mídia diretamente após cada gravação.
- Rebobine e verifique esta gravação imediatamente. (Em caso de problemas; é o microfone externo. ligado? Verifique todos os cabos, fiação, configurações, proteções, etc.)

#### 2). Gravando vídeo

- Se você precisar gravar ao ar livre, tente encontrar um lugar calmo e protegido ao vento.
- Se necessário, use protetor de vento para o (s) microfone (s).
- Evite gravar quando o fundo estiver muito brilhante. Filmar em direção ao sol trará uma câmera de vídeo ao seu limite. Tente encontrar outro lugar e filmar 'com' a luz do sol.
- Não exponha o equipamento à luz solar por muito tempo, pois os dispositivos podem esquentar para que as especificações fornecidas não sejam mais válidas. Tente encontrar um lugar sombreado.
- Coloque o microfone o mais próximo possível, mas não muito próximo do ponto de interesse.
- Esteja ciente das características do microfone. Verifique se o microfone está ajustado corretamente.
- Não use cabos de extensão longa. Mantenha o cabo do microfone externo o mais curto possível.
- Anexe a câmera a um tripé. Se isso não for possível, use a função "Steady Shot" para obter uma imagem estável. (consulte o manual do dispositivo)
- Use o modo SP (jogo padrão). A gravação de áudio será definida como 48 KHz, 16Bit. (Isso garante a gravação linear do PCM)
- Nunca use LP (modo de reprodução longa). Isso resultará em uma gravação de áudio não linear (32 kHz, 12 bits)
- Antes de começar a filmar, faça uma calibração do balanço de brancos, toda vez que você alternar entre as filmagens internas e externas. (consulte o manual do dispositivo)
- Não registre dados relevantes nos primeiros 30 segundos e nos últimos 30 segundos de uma fita. Isso levará a problemas durante o processo de digitalização.

<sup>177</sup> Figura retirada e adaptada de The Language Archive (MPI for Psycholinguistics). Disponível em: <<https://www.mpi.nl/corpus>>. Acesso em: 16 set. 2021.

**ANEXO E - Guia DOBES para gravação de vídeo - parte 2<sup>178</sup>****Fazendo gravações de vídeo**

## Comentários, dicas e dicas para gravar vídeo

- Não registre dados relevantes nos primeiros 30 segundos e nos últimos 30 segundos de uma fita. Isso levará a problemas durante o processo de digitalização.
- Verifique se há um sinal de vídeo constante gravado na fita. Não produza lacunas entre as diferentes cenas. Uma nova cena deve começar diretamente no final da anterior. Isso pode ser alcançado com a ajuda do botão END SEARCH ou outras funções disponíveis na maioria das filmadoras (consulte o manual do dispositivo).
- Não tire fotos e guarde-as na fita, use o cartão de memória extra.
- Não deixe a câmera de vídeo no modo de espera por mais tempo do que o necessário. Alguns dispositivos serão trocados após um certo tempo automaticamente. (consulte o manual do dispositivo)
- Esteja ciente dos problemas que podem ocorrer ao usar a função AUTO FOCUS. Mover objetos na cena pode acionar o circuito de foco e causar problemas de nitidez; portanto, às vezes é melhor usar o MANUAL FOCUS
- As filmagens internas só podem ser possíveis com uma lente grande angular extra para cobrir toda a cena.
- Após cada gravação, verifique a fita imediatamente. Em caso de problemas, execute a fita de limpeza.
- Você pode copiar as faixas de áudio da câmera de vídeo usando o conector de saída de linha e armazenar o áudio em outras mídias de áudio para transcrever. Com algumas filmadoras, isso só é possível com a ajuda de cabos especiais.
- Armazene as fitas gravadas em local seco e fresco e defina a guia de salvamento de registros para evitar mais gravações.

<sup>178</sup> Figura retirada e adaptada de The Language Archive (MPI for Psycholinguistics). Disponível em: <<https://www.mpi.nl/corpus>>. Acesso em: 16 set. 2021.

## ANEXO F - Projetos de documentação DOBES em diversas partes do mundo<sup>179</sup>

### PROJETOS DE DOCUMENTAÇÃO

Dentro do programa DOBES, os seguintes projetos de documentação estão documentando línguas altamente ameaçadas de extinção:

#### AMÉRICA DO NORTE E MESOAMÉRICA

- Lakota
- Wichita (corpus)
- Lacandon (corpus)
- Beaver (corpus)
- Hooçak (corpus)
- Déljñe (corpus)
- Chontal (corpus)

#### AMÉRICA DO SUL

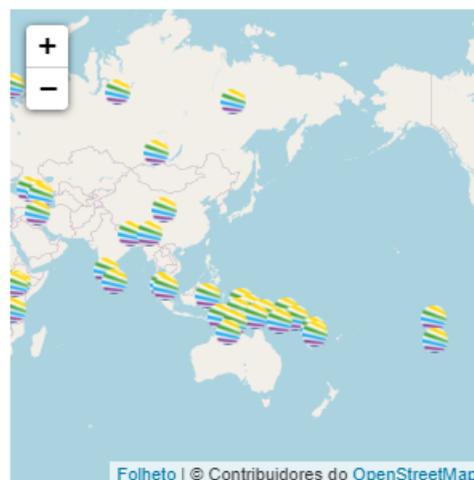
- Chipaya (corpus)
- Uru-Chipaya (corpus)
- Tsafiki (corpus)
- Trumai (corpus)
- Pessoas do Centro (corpus)
- Movima (corpus)
- Aché (corpus)
- Kuikuro (corpus)
- Yurakaré (corpus)
- Chaco Languages (corpus)
- Southeastern Rondônia (corpus)
- Cashinahua (corpus)
- Awetí (corpus)
- Baure (corpus)

#### EUROPA E ÁSIA

- Chintang/Puma (corpus)
- Minderico (corpus)
- Salar/Monguor (corpus)
- Shiri / Sanzhi (corpus)
- Lower Sorbian (corpus)
- Sri Lanka Malay (corpus)
- Svan / Udi / Tsova-Tush (corpus)
- Tangsa / Tai / Singpho (corpus)
- Kurumba Languages (corpus)
- Kola Sámi (corpus)
- Khinalug (corpus)
- Tofa (corpus)
- Gorani (corpus)
- Éven (corpus)
- Nenets (corpus)

#### ÁFRICA

- Taa (corpus)
- Tima (corpus)
- Akie (corpus)
- Bainouk (corpus)
- Oyda (corpus)
- Nsambaan / Nsong / Ngong (corpus)
- Bagyeli/Bakola (corpus)
- Bayso / Haro (corpus)
- Bezen (corpus)
- Isubu / Wovia (corpus)
- Kyanga/Shanga (corpus)
- Laal (corpus)
- ≠ Akhoe Hai // om (corpus)



#### SUDESTE ASIÁTICO E OCEANIA

- Marquesan (corpus)
- West Ambrym (corpus)
- Wooi (corpus)
- CELD Papua (corpus)
- Waima'a (corpus)
- Vurès/Vera'a (corpus)
- MEL project / Tuamotuan (corpus)
- Ujir (corpus)
- Saliba / Logea (corpus)
- Semaq Beri / Batek (corpus)
- Totoli (corpus)
- Iwaidja (corpus)
- Jaminjungan e Eastern Ngumpin (corpus)
- Morehead (Nen / Kómnzó) (corpus)
- Bena Bena (corpus)
- Savosavo (corpus)
- Semang (corpus)
- Teop (corpus)

<sup>179</sup> Figura retirada e adaptada de Projetos de documentação. Disponível em: <<https://dobes.mpi.nl/projects/?lang=pt>>. Acesso em: 12 nov. 2021.

## ANEXO G - Padrões de formato e codificação DOBES - parte 1<sup>180</sup>

### Acordos de formato e codificação DOBES

Este guia descreve os Acordos DOBES em relação ao formato de arquivo e aos padrões de codificação, conforme definidos na fase piloto e emergidos lentamente devido a avanços tecnológicos.

#### Quadro geral:

- O Arquivo de Idiomas DOBES deve atender a dois propósitos principais: (1) fornecer acesso direto aos dados da documentação onde o acesso é permitido e (2) preservar o material para as gerações futuras.
- O programa DOBES é um projeto que compreende muitas equipes de documentação que entregam (cópias) todo o material ao arquivo DOBES, a fim de criar um repositório coerente de recursos de linguagem.

#### 1. Observações gerais

Temos que distinguir entre **formatos de arquivo e codificações (F + E)** e aqueles que são oferecidos pelos doadores. O arquivo F + E deve aderir a padrões abertos, enquanto o F + E oferecido pelos doadores está relacionado a acordos de fluxo de trabalho entre doadores e arquivistas. Eles dependem do tipo de manipulação e conversões que o arquivista é capaz de lidar. Naturalmente, o conjunto de F + E oferecido é um superconjunto do arquivo F + E. Também temos que distinguir entre **representação e apresentação F + E**. A representação F + E é relevante para o armazenamento neutro e persistente de dados, enquanto a apresentação F + E determina como os dados são apresentados aos usuários. Nesta nota, nos preocuparemos principalmente com os formatos de representação, uma vez que a apresentação F + E pode ser derivada dos back-end.

Em geral, **codificações** descreva como informações como caracteres ou fluxos de vídeo são representadas como padrões de bits. **Formatos** descreva como esses padrões são empacotados em objetos de arquivo, ou seja, arquivos.

Como a tecnologia está mudando e, portanto, os acordos também mudarão com o tempo, nós o fazemos ser explícito sobre a data em que este documento foi criado : **Abril de 2005**

#### 2). Codificações de arquivo

As seguintes codificações são aceitas para o arquivo :

- Para codificação de caracteres, UNICODE é o padrão, ou seja, todo o material de entrada deve ser transformado em UNICODE e as pessoas precisam criar o UNICODE sempre que possível.
- Para catalogação, o padrão IMDI é obrigatório (XML mais esquema IMDI).
- O material textual estruturado (notações e lexica) deve ser estruturado com XML e eles devem aderir a esquemas genéricos. Neste momento, o arquivo DOBES usa EAF (ELAN Annotation Format) e LMF (ISO Lexical Markup Framework) como esquemas subjacentes.
- Para textos não estruturados, o arquivo aceitará texto sem formatação e HTML. Em casos excepcionais, os documentos PDF também precisam ser aceitos.
- Para representações de áudio digital, o PCM linear é o padrão. Em geral, 16 bits (96 dB) e 44,1 / 48 kHz são suficientes, no entanto, 24 bits (144 dB) e 96 kHz também são aceitos. (MiniDisc e MP3 não são aceitos como codificações de back-end de arquivo!)
- Para imagens digitais, JPEG, PNG e TIFF são os padrões aceitos. No futuro, esperamos cada vez mais RAW / TIFF como formatos de representação, uma vez que o JPEG implica compactação com perda.
- Para vídeo digital, o MPEG2 é o padrão de arquivo atual. Alguns recursos antigos podem estar no MPEG1, que deve ser aceito. As informações de áudio incluídas nos fluxos de vídeo também devem ser extraídas como dados lineares do PCM.
- Para vídeo, as normas PAL e NTSC são aceitas.

Como sabemos que as conversões geralmente não estão livres da perda de informações, também armazenaremos alguns arquivos originais no arquivo, como o Shoebox. Para fins de apresentação, geraremos arquivos MP3 e MPEG4 a partir dos formatos de back-end mencionados acima.

Peter Wittenburg, MPI para Psicolinguística, 3.4.2005

<sup>180</sup> Figura retirada e adaptada de Acordos de formato e codificação DOBES. Disponível em: <<https://dobes.mpi.nl/dobesprogramme/?lang=pt>>. Acesso em: 16 nov. 2021.

## ANEXO H - Padrões de formato e codificação DOBES – parte 2<sup>181</sup>

### Acordos de formato e codificação DOBES 2

#### 3). Formatos de arquivo

Para algumas codificações, os formatos são predefinidos pelo padrão de codificação escolhido. Os seguintes formatos são aceitos para o arquivo morto :

- Para textos estruturados, arquivos XML validados são aceitos (veja acima).
- Para textos não estruturados, arquivos de texto sem formatação ou HTML são aceitos. Exceções em PDF são possíveis.
- Para fluxos de áudio, a embalagem do arquivo WAV é o padrão.
- Para imagens, é aceita a embalagem de arquivos JPEG, PNG ou TIFF.
- Para vídeos, todos os formatos são convertidos em arquivos formatados em MPG.

#### 4). Entrada F + E

Este capítulo pode estar sujeito a alterações, ou seja, ninguém pode confiar na disponibilidade dos serviços mencionados. Em muitos casos, a conversão não é trivial, ou seja, é altamente recomendável que sejam feitos acordos iniciais de fluxo de trabalho. Além do F + E mencionado acima, os seguintes F + E são aceitos como doador F + E :

- Anotação Shoebox / Toolbox, léxico e arquivos auxiliares, como tipo de arquivo etc; recomendamos o uso do Toolbox para fornecer Unicode
- Arquivos de anotação de transcritor
- Arquivos de anotação e léxico CHAT com a recomendação de usar a codificação Unicode
- Gravações MiniDisc (embora não sejam recomendadas devido a filtragem severa)
- DV como codificação de vídeo nativa criada pela maioria das câmeras atuais
- vários formatos de mídia, como AVI, MOV, MPV, MPA

No passado, o arquivo DOBES também processava outros F + E, mas isso precisa ser negociado entre as equipes e o arquivista, pois algumas das conversões incluídas são muito demoradas.

- Normas de operadora: MiniDisc, Cassette, Reel-to-Reel, Uher4400, Hi-8, UMatic, VCD
- Tipos de documento: arquivos WORD com marcação de estrutura clara e consistente
- XML: outros arquivos XML com marcação consistente, mesmo que nenhum esquema seja fornecido
- Codificação / fontes de caracteres: SIL IPA Sets, IPA Kiel

#### 5). Ferramentas recomendadas

As ferramentas não fazem parte dos acordos, embora possamos dar uma série de recomendações para ferramentas a partir das quais sabemos que o F + E acordado pode ser gerado. Esta lista não pode ser vista como completa, pois pode haver mais ferramentas que suportam o F + E mencionado. Em particular, não faz sentido listar todas as ferramentas de mídia úteis.

- Caixa de som / caixa de ferramentas: para criar formatos Shoebox para léxico e anotações, o Toolbox é altamente recomendado, pois permite trabalhar com o Unicode
- Transcriber: para criar arquivos Transcriber
- Ferramentas CLAN: para criar arquivos formatados por bate-papo; recomendamos o uso da versão Unicode
- ELAN: criar arquivos EAF e importar / exportar arquivos de anotação Shoebox, Chat e Transcriber
- Onto-ELAN: para criar arquivos EAF usando ontologias
- Editor IMDI: para criar arquivos IMDI validantes

LEXUS: para criar arquivos LMF e importar / exportar arquivos de léxico Shoebox e Chat  
Muitas vezes, neste texto, falamos sobre o UNICODE. Principalmente, o esquema de codificação UTF-8 flexível é aplicado pelo software.

<sup>181</sup> Figura retirada e adaptada de Acordos de formato e codificação DOBES. Disponível em: <<https://dobes.mpi.nl/dobesprogramme/?lang=pt>>. Acesso em: 16 nov. 2021.

**ANEXO I - Relação de projetos apoiados pelo DOBES e ELDP no Brasil<sup>182</sup>**

<b>Língua (família)</b>	<b>Linguista</b>	<b>Instituição (s) 1</b>
<i>DOBES:</i>		
Kuikuro (Carib)	Franchetto	Museu Nacional
Trumai (isolado)	Guirardello	MPI Nijmegen / Museu Goeldi Universidade Livre de Berlim / Museu Goeldi
Aweti (Tupí)	Drude	Goeldi
Kaxuyana (Carib)	Meira	Leiden / Museu Goeldi
Bakairi (Carib)	Meira	Leiden / Museu Goeldi
Mawé (Tupí)	Meira	Leiden / Museu Goeldi
Kaxinawá	Camargo	CNRS
Aikanã (isolado)	van der Voort	MPI Nijmegen / Museu Goeldi
Kwazá (isolado)	van der Voort	MPI Nijmegen / Museu Goeldi
<i>ELDP:</i>		
Puruborá (Tupí)	Galucio	Museu Goeldi
Sakurabiat (Tupí)	Galucio	Museu Goeldi
Karo (Tupí)	Gabas	Museu Goeldi Universidade Livre de Bruxelas / Univ de São Paulo
Ayuru (Tupí)	Demolin	Paulo
Salamãý (Tupí)	Moore	Museu Goeldi
Xipayá (Tupí)	Rodrigues	Universidade Federal do Pará
Apurinã (Arawak)	Facundes	Universidade Federal do Pará Universidade de Chicago / Univ Federal de Goiás
Ofayé (Macro-Jê)	Ribeiro	Goiás
Kaduwéu (Guaykuru)	Sandalo	Universidade Estadual de Campinas
Enawenê Nawé (Arawak)	de Resende	Museu Nacional Universidade Radboud Nijmegen / Museu Goeldi
Oro Win (Chapakura)	Birchall	
Waikhana (Tukano Oriental)	Stenzel	Univ Federal do Rio de Janeiro
Wanano (Tukano Oriental)	Stenzel	Univ Federal do Rio de Janeiro
Kanamari (Katukina)	Dienst	Universidade Goethe
Akuntsu (Tupí)	Aragão	Universidade de Utah
Kubeo (Tukano)	Chacon	Universidade de Utah
Desana (Tukano)	Silva	Universidade de Utah
Gavião (Tupí)	Meyer	Museu Goeldi
Suruí (Tupí)	Meyer	Museu Goeldi
Dâw (Nadahup)	Epps	Universidade do Texas, Austin
Paressi-Haliti (Arawak)	da Silva	Universidade Federal do Rio de Janeiro
Paressi-Haliti (Arawak)	Brandão	Universidade do Texas, Austin

<sup>182</sup> Quadro retirado e adaptado de Báez e Galucio (2016, p. 32)

## ANEXO J - Línguas indígenas documentadas no MPEG e armazenadas no TLA (DOBES/Max Planck)<sup>183</sup>



The screenshot shows the TLA interface. On the left is a hierarchical tree of language corpora. The root is 'ALIM', which branches into '350 Línguas - Vocabulários Seletos' and 'A-Z'. Under 'A-Z', there are numerous language-specific folders, each with a colored icon indicating its access level. The right side of the interface displays the 'Acervo de Línguas Indígenas - Museu Paraense Emílio Goeldi' logo and a list of functions available for the selected node.

**ALIM**

- 350 Línguas - Vocabulários Seletos
  - ? 350-Línguas-VocabularioSeletos.pdf
  - Vocabulários Seletos
- A-Z
  - Aikana-TBA
  - Atividades-indef
  - Atividades-rituais
  - Elicitacao-gramatica-fonologia
  - Elicitacao-lexico
  - Eventos-de-fala-nao-tradicionais
  - Eventos-de-fala-tradicionais
  - Musica-instrumental
  - Musica-vocal
  - Outros-materiais
  - Apalai-APY
  - Apurina-APU
  - AraraDoPara-AAP
  - Arikapu-ARK
  - Arua-ARX
  - Asurini-ASU
  - Bakairi-BKQ
  - Banawa-JAA
  - Baniwa-BWI
  - CintaLarga-CIN
  - Daw-KWA
  - Desana-DES
  - Djeoromitxi-JBT
  - Dobes-ESR
  - GaviaoDeRondonia-GVO
  - Guaja-GVJ
  - GuaraniKaiowa-KGK
  - Hixkaryana-HIX
  - Hupda-JUP
  - Ingariko-AKE
  - Juruna-JUR
  - Kanoe-KXO
  - Karaja-KPJ
  - ? Karipuna-KUQ
  - KaripunaDoAmapa-KMV
  - Karitiana-KTN

**The Language Archive**  
MAX PLANCK INSTITUTE FOR PSYCHOLOGICAL SCIENCES

**MUSEU GOELDI**  
Coordenação de Ciências Humanas  
**COCHS**

### Acervo de Línguas Indígenas - Museu Paraense Emílio Goeldi

Please see below for a brief explanation on how to use the archive browser or get the full manual [here](#).

To contact us, send a e-mail to: [linguistica@museu-goeldi.br](mailto:linguistica@museu-goeldi.br) (no spaces).

**Brief explanation:**

On the left, you see a hierarchical organisation of the various corpora that are archived at TLA. You can browse through these corpora by opening up the various branches. When you select a certain node in the hierarchy, a number of buttons become available on the right side for various functions, depending on the type of node you have selected. These functions can be:

- View:** View an archived resource. \*
- Download:** Download an archived resource to your own computer. \*
- Download All:** Download all archived resources that belong to the selected resource bundle (session).
- Metadata Search:** Perform a search in all metadata descriptions underneath the selected branch.
- Content Search:** Perform a search in all textual archived resources below the selected branch. \*
- Citation:** Get a persistent link for citing the resource or branch in a publication or for linking to it on a web site.
- Manage Access:** Define access permissions on the selected branch (requires management permissions).
- Request Access:** Request access to archived materials underneath the selected branch.
- Version Info:** View information about different versions of the selected resource.

\* these functions work on resources to which you have access.

There are 4 access levels for archived materials, indicated by 4 coloured icons:

- Open:** No login or registration is required
- Available to registered users:** You can register as a user of the archive by clicking on the "register" link at the top.
- Access needs to be requested:** By using the "Request Access" function, you can apply for access to these materials as a registered user. The depositor will be asked whether or not access can be granted to you, which may take some time.
- Closed:** These materials are currently not accessible, mostly due to the sensitivity of the material

Materials that are located in other archives have an icon with no color ○ Their access level can vary

In addition to these access levels, it may be required to agree to a license agreement.

<sup>183</sup> Figura retirada e adaptada de Báez e Galucio (2016, p. 32)

## ANEXO K - Conteúdo parcial das coleções brasileiras no DMA – parte 1<sup>184</sup>

# Search

This is a view of all the collections in the Digital Media Archive. You can search for keywords or filter by the categories at right.

Now showing 1 - 10 of 15 collections.

Search
Clear

Com licença! Brazilian Portuguese for Spanish speakers
By Request

Portuguese
Spanish

Portuguese
Spanish
English
Brazil
South America

**Collection Descriptive Notes:**  
1) Introduction, Unit 1;2) Units 2-3;3) Spoken vocabulary A-Z

**Collection Technical Notes:**

**Contributor:**

Musica indígena Brasileña
By Request

No linguistic content
Brazil
South America

**Collection Descriptive Notes:**

**Collection Technical Notes:** From the McQuown donation

**Contributor:** Unnamed

Português do Brasil para estrangeiros;; conversação, cultura e criatividade
By Request

Portuguese

Portuguese

Portuguese
Brazil
South America

**Collection Descriptive Notes:** Lessons 1-15

**Collection Technical Notes:**  
From the McQuown donation. The model numbers of the cassettes are MK-7 PBE -1 & -2.

**Contributor:**

### Filter by languages

- Portuguese (12)
- English (3)
- Krenak (2)
- Tenharim (2)
- Kamayurá (1)
- Kayapó (1)
- No linguistic content (1)
- Spanish (1)

### Filter by subject languages:

- Portuguese (6)
- Kayapó (1)
- Krenak (1)
- Tenharim (1)

### Filter by locations

- Brazil
- South America (15)

### Filter by access

- By Request (12)
- Login Required (3)

### Filter by contributor

- (8)
- Unnamed (4)
- Akö, consultant (1)
- Anonymous consultants (1)
- de Andrade, Carlos Drummond (1)

<sup>184</sup> Figura retirada e adaptada de DMA. Disponível em: <[https://dma.uchicago.edu/search?f\[0\]=dma\\_collection\\_locations%3A7663](https://dma.uchicago.edu/search?f[0]=dma_collection_locations%3A7663)>. Acesso em 20 dez. 2021.

ANEXO L - Conteúdo parcial de coleções brasileiras no DMA – parte 2<sup>185</sup>**Contributor:**

Interviews in Krenak (Borun) and Kamaiuré (/collections/447)

By Request (/access/by-request)

Krenak (/languages/krenak)

Portuguese (/languages/portuguese)

Krenak (/languages/krenak) Kamayurá (/languages/kamayur-) Portuguese (/languages/portuguese)

Brazil (/locations/brazil) South America (/locations/south-america)

**Collection Descriptive Notes:****Collection Technical Notes:** (Unreviewed) WAVE file is now converted to an MP3.**Contributor:** Unnamed

An anthology of the poetry of Carlos Drummond de Andrade (/collections/540)

By Request (/access/by-request)

Portuguese (/languages/portuguese)

Portuguese (/languages/portuguese) Brazil (/locations/brazil) South America (/locations/south-america)

**Collection Descriptive Notes:**

**Band One:** Disc 1, Side 1: 1) *Infância*, 2) *No meio do caminho*, 3) *Confidência de itabirano*, 4) *Quadrilha*, 5) *Os ombros supportam o mundo*, 6) *Mãos dadas*, 7) *Mundo grande*, 8) *José*, 9) *Viagem na família*, 10) *Procura da poesia*; Disc 1, Side 2: 1) *O mito*, 2) *O Lutador*, 3) *Memória*, 4) *Morte do leiteiro*, 5) *Confissão*, 6) *Consolo na praia*, 7) *Oficina irritada*; Disc 2 side 1: 1) *Fazenda*, 2) *Caso do vestido*, 3) *Estrambote melancólico*, 4) *O enterrado vivo*, 5) *Destruição*, 6) *Intimação*, 7) *Alta cirurgia*, 8) *Para sempre*, 9) *Canto do Rio em sol*, 10) *Boitempo*; Disc 2, Side 2: 1) *Cantiguinha*, 2) *Os pacifistas*, 3) *Cultura francesa*, 4) *Falta um disco*, 5) *Amor e seu tempo*, 6) *Obrigado*, 7) *Lira Romantiquinha*, 8) *O Homem, as viagens*, 9) *Essas coisas*, 10) *Parolagem da vida*, 11) *Declaração de amor*

**Collection Technical Notes:****Contributor:** de Andrade, Carlos Drummond

Music from Brazil (/collections/915)

By Request (/access/by-request)

Portuguese (/languages/portuguese)

Portuguese (/languages/portuguese) Brazil (/locations/brazil) South America (/locations/south-america)

**Collection Descriptive Notes:**

From the notes on the back of the box: "Musica indígena brasileira", "Português alemão de Curitiba" [sp?], and Chico Buange [sp?] canta lo suyo"

**Collection Technical Notes:** (Gift of Norman A. McQuown to the LLA in April 2004)**Contributor:** Unnamed

Kayapó sound tracks (Turner, c. 1979) (/collections/1061)

By Request (/access/by-request)

Kayapó (/languages/kayap-)

Kayapó (/languages/kayap-) Brazil (/locations/brazil) South America (/locations/south-america)

**Collection Descriptive Notes:** Terry Turner interviews various native speakers (male and female) in Brazil.**Collection Technical Notes:**

Master is the copy made in the LLA from the off-dub (3 ¼ ips, 5" reel, 1-mil stock), made by the BBC from Turner's documentary film on the Amazonian Indians. The WAVE files are converted to MP3s.

**Contributor:**

Two Krenakôre myths (Schwartzman, 1981) (/collections/1948)

By Request (/access/by-request)

Krenak (/languages/krenak)

Krenak (/languages/krenak) Brazil (/locations/brazil) South America (/locations/south-america)

<sup>185</sup> Figura retirada e adaptada de DMA. Disponível em: <[https://dma.uchicago.edu/search?f\[0\]=dma\\_collection\\_locations%3A7663](https://dma.uchicago.edu/search?f[0]=dma_collection_locations%3A7663)>. Acesso em 20 dez. 2021.

ANEXO M - Resultado parcial de busca por "Brazil" no AILLA<sup>186</sup>


The Archive of the Indigenous Languages of Latin America

---

Collections Languages Countries Announcements Organizations Persons

---

Home » brazil (x) » Search results

## Search results

(1 - 50 of 1,240)

1 2 3 4 5 ... next > last »

	<b>Label:</b> Brazil
	<b>Label:</b> Review of "The Elusive Eden: Frank McMullan's Confederate Colony in Brazil" <b>Title:</b> Review of "The Elusive Eden: Frank McMullan's Confederate Colony in Brazil" <b>Título:</b> Reseña de libro "El Edén ilusorio: La colonia de confederados de Frank McMullan e" <b>Language(s):</b> Multiple <b>Parent Collection:</b> Amazonian Languages Collection of Arthur P. Sorensen, Jr.
	<b>Label:</b> Häwäg and b'atib : The Balance between Health and Disease among the Hupd'äh in the Upper Rio Negro Region, Brazil <b>Title:</b> Häwäg and b'atib : The Balance between Health and Disease among the Hupd'äh in the Upper Rio Negro Region, Brazil <b>Título:</b> Häwäg y b'atib : El equilibrio entre salud y enfermedad entre los hupdè del Alto Río Negro, Brasil <b>Language(s):</b> Hup
	<b>Label:</b> Jaime_on_priest_in_Brazil.wav <b>Language(s):</b> Barasana
	<b>Label:</b> Dreams Interpretation <b>Title:</b> Dreams Interpretation <b>Título:</b> Los sueños y sus interpretaciones entre los indígenas macú hupdah de Brasil <b>Language(s):</b> Hup
	<b>Label:</b> dreams-interpretation.pdf <b>Language(s):</b> English, Hup
	<b>Label:</b> Nadéb Collection <b>Title:</b> Nadéb Collection <b>Título:</b> Colección Nadéb <b>Language(s):</b> Nadéb

### User

- My Account
- Log Out

### Main Menu

- Welcome
- Browse
- Depositor Info
- Rights
- Events
- AILLA Papers
- Contact
- Contribute
- News and Announcements
- FAQ

<sup>186</sup> Figura retirada e adaptada de AILLA. Disponível em: <<https://ailla.utexas.org/islandora/search/brazil?type=dismax>>. Acesso em 31 jan. 2022.

## ANEXO N - Resultado parcial de busca por “Mbyá Guarani” no AILLA<sup>187</sup>


The Archive of the Indigenous Languages of Latin America

[Collections](#)
[Languages](#)
[Countries](#)
[Announcements](#)
[Organizations](#)
[Persons](#)
Search Repository

Home » guarani mbya<sup>03</sup> » Search results

### Search results

(1 - 50 of 224)

1 2 3 4 5 next > last »

	<p><b>Label:</b> Dependency Treebank derived from the Mbyá Guarani collection of Robert Dooley</p> <p><b>Title:</b> Dependency Treebank derived from the Mbyá Guarani collection of Robert Dooley</p> <p><b>Titulo:</b> Corpus sintácticamente anotado derivado de la colección Mbyá Guarani de Robert Dooley</p> <p><b>Language(s):</b> Guaraní, Mbyá</p>
	<p><b>Label:</b> Guaraní, Mbyá</p>
	<p><b>Label:</b> Mbyá Guarani Collection of Robert Dooley</p> <p><b>Title:</b> Mbyá Guarani Collection of Robert Dooley</p> <p><b>Titulo:</b> Colección Mbyá Guarani de Robert Dooley</p>
	<p><b>Label:</b> Dependency Treebank derived from the Mbyá Guarani collection of Robert Dooley, version 0</p> <p><b>Title:</b> Dependency Treebank derived from the Mbyá Guarani collection of Robert Dooley, version 0</p> <p><b>Titulo:</b> Corpus sintácticamente anotado derivado de la colección Mbyá Guarani de Robert Dooley</p> <p><b>Language(s):</b> Guaraní, Mbyá</p>
	<p><b>Label:</b> gun_dooley-ud-v0.conllu</p>
	<p><b>Label:</b> The giant</p> <p><b>Title:</b> The giant</p> <p><b>Titulo:</b> O gigante</p> <p><b>Language(s):</b> Guaraní, Mbyá</p> <p><b>Parent Collection:</b> Mbyá Guarani Collection of Robert Dooley</p>
	<p><b>Label:</b> The poor man</p> <p><b>Title:</b> The poor man</p> <p><b>Titulo:</b> História do homem pobre</p> <p><b>Language(s):</b> Guaraní, Mbyá</p> <p><b>Parent Collection:</b> Mbyá Guarani Collection of Robert Dooley</p>
	<p><b>Label:</b> gun_dooley-ud-v1.conllu</p> <p><b>Language(s):</b> Guaraní, Mbyá</p>
	<p><b>Label:</b> Dependency Treebank derived from the Mbyá Guarani collection of Robert Dooley, version 1</p> <p><b>Title:</b> Dependency Treebank derived from the Mbyá Guarani collection of Robert Dooley, version 1</p> <p><b>Language(s):</b> Guaraní, Mbyá</p>
	<p><b>Label:</b> When salt and soap were unknown</p> <p><b>Title:</b> When salt and soap were unknown</p> <p><b>Titulo:</b> Quando o sal e o sabão eram desconhecidos</p> <p><b>Language(s):</b> Guaraní, Mbyá</p> <p><b>Parent Collection:</b> Mbyá Guarani Collection of Robert Dooley</p>

**User**

- My Account
- Log Out

**Main Menu**

- Welcome
- Browse
- Depositor Info
- Rights
- Events
- AILLA Papers
- Contact
- Contribute
- News and Announcements
- FAQ

<sup>187</sup> Figura retirada e adaptada de AILLA. Disponível em: <https://ailla.utexas.org/islandora/search/guarani%20mbya?type=dismax>. Acesso em 31 jan. 2022.

## ANEXO O – Coleção de língua indígena armazenada no AILLA<sup>188</sup>

### Colección: Tepehua Collection

Visualizar los recursos en esta colección

<b>Título</b>	
<b>Título español</b>	Colección Tepehua
<b>Título inglés</b>	Tepehua Collection
<b>Lenguas en esta colección:</b>	Tepehua de Huehuetla
<b>Coleccionista(s)</b>	Susan Kung
<b>Depositante(s)</b>	Susan Kung
<b>Sitio del Proyecto/Coleccionista</b>	
<b>Descripción</b>	Esta colección de grabaciones de audio incluye una variedad de generos de discurso. Hay grabaciones de traducciones al español para la mayor parte de estos recursos.
<b>Referencias</b>	

### Resumen de los contenidos de la colección

<b>Géneros</b>	Narrativa; Conversación; Entrevista; Lista de palabras; Elicitación; Música instrumental; Descripción; Juego de datos; Historia; Etnografía; Canción; Receta; Procedimiento; Gramática; Tesis; Léxico; Esbozo; Imágen; Fotografía		
<b>Número de archivos archivales</b>	276	<b>Porcentaje de archivos restringidos</b>	1
<b>Número de grabaciones de audio</b>	163	<b>Duración total de audio</b>	17:22:44
<b>Número de grabaciones de video</b>	95	<b>Duración total de video</b>	23:47:33
<b>Número de textos digitales*</b>	11	<b>Páginas de textos digitales</b>	1231
<b>Páginas de textos en manuscrito</b>	4	<b>Número de imágenes</b>	3
<b>Memoria para objetos archivales</b>	66.9G		
<b>Porcentaje de recursos que incluyen transcripciones</b>			2

### 25 **Ver esta colección:**

Susan. "Colección Tepehua" El Archivo de los Idiomas Indígenas de Latinoamérica: [www.ailla.utexas.org](http://www.ailla.utexas.org). Media: audio, video, text, image. Acceso: 1% restringido.

<sup>188</sup> Figura retirada e adaptada de AILLA. Disponível em: <[https://ailla.utexas.org/sites/default/files/documents/como\\_usar\\_AILLA\\_2.pdf](https://ailla.utexas.org/sites/default/files/documents/como_usar_AILLA_2.pdf)>. Acesso em 31 jan. 2022