SciELO Preprints

Estado da publicação: O preprint foi publicado em outro meio.
DOI do preprint publicado: https://doi.org/10.25189/rabralin.v24i2.2384

DICIONÁRIOS MULTIMÍDIA: UMA TECNOLOGIA SOCIAL PARA AUXILIAR A APRENDIZAGEM E REVITALIZAÇÃO DE LÍNGUAS MINORIZADAS

Ana Vilacy Galucio, Joshua Birchall, Ivan Rocha, Saulo Rodrigo Teixeira Brito, Matheus Augusto Ribeiro Soares, Israel Nathan Marcon, Douglas da Costa Rodrigues Junior, Ellison dos santos, Victor Siqueira Rocha, Antônia Fernanda de souza nogueira, Flávio Henrique Ferreira Pinheiro, Juliana Ketelen Souza Solano, Bruno Pinto Silva

https://doi.org/10.1590/SciELOPreprints.13693

Submetido em: 2025-10-08
Postado em: 2025-10-10 (versão 1)
(AAAA-MM-DD)

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.13693

DICIONÁRIOS MULTIMÍDIA: UMA TECNOLOGIA SOCIAL PARA AUXILIAR A APRENDIZAGEM E REVITALIZAÇÃO DE LÍNGUAS MINORIZADAS

MULTIMEDIA DICTIONARIES: A SOCIAL TECHNOLOGY TO SUPPORT THE LEARNING AND REVITALIZATION OF MINORITY LANGUAGES

Ana Vilacy Galúcio
Museu Paraense Emílio Goeldi. https://orcid.org/0000-0003-0168-1904.
Joshua Birchall
University of New Mexico. https://orcid.org/0000-0002-6110-6904.
Ivan Rocha
Museu Paraense Emílio Goeldi. https://orcid.org/0000-0001-6533-4528
Saulo Rodrigo Teixeira Brito
Museu Paraense Emílio Goeldi. https://orcid.org/0009-0004-6425-9641.
Matheus Augusto Ribeiro Soares
Universidade Federal do Pará. https://orcid.org/0009-0008-1506-9007.
Israel Nathan Marcon
Universidade do Estado do Pará. https://orcid.org/0009-0002-4117-2872.
Douglas da Costa Rodrigues Junior
Universidade Federal do Pará. https://orcid.org/0009-0006-2165-0653.
Ellison Santos
Museu Paraense Emílio Goeldi. https://orcid.org/0009-0002-6992-8985.
Victor Siqueira Rocha
Universidade Federal do Pará. Museu Paraense Emílio Goeldi. https://orcid.org/0009-0002-4335-9024.
Antônia Fernanda de Souza Nogueira
Universidade Federal do Pará. https://orcid.org/0000-0002-3105-6520.
Flávio Henrique Ferreira Pinheiro
Universidade Federal do Pará. Museu Paraense Emílio Goeldi. https://orcid.org/0009-0000-2452-945X.
Juliana Ketelen Souza Solano
Universidade Federal do Pará. Museu Paraense Emílio Goeldi. https://orcid.org/0009-0008-2754-1460.
Bruno Pinto Silva
University of New Mexico. https://orcid.org/0000-0002-9933-3037.

RESUMO

Este artigo aborda a importância dos dicionários multimídia como um produto multifuncional e multiuso desenvolvidos como uma tecnologia social voltada para a aprendizagem de línguas ameaçadas. Os dicionários multimídia para línguas indígenas constituem uma metodologia reaplicável, de uso livre e fácil acesso, que pode ser usada pelas comunidades em larga escala e sem necessidade de acesso à internet. O uso de tecnologia como smartphones, tablets ou computadores, ainda que incipiente, é uma realidade em muitas comunidades indígenas, o que favorece a implementação de dicionários em formatos digitais, cujas entradas lexicais podem ser ilustradas com recursos multimídias. O artigo apresenta inicialmente o conceito, a metodologia e a tecnologia envolvidas na elaboração dos dicionários e descreve as macro e microestruturas adotadas nessas obras, decididas conforme a necessidade dos falantes ou potenciais falantes da língua. O trabalho discute ainda o papel dos dicionários multimídia no planejamento linguístico em práticas de revitalização de línguas minorizadas e ameaçadas, a partir da experiência de elaboração de dicionários de oito línguas indígenas em desenvolvimento no setor de linguística do Museu Paraense Emílio Goeldi (MPEG) em

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.13693

colaboração com a Universidade do Novo México (UNM, EUA): Kanoé, Sakurabiat, Oro Win, Djeoromitxi, Puruborá, Makurap, Wayoro e Salamãi. No contexto de vulnerabilidade e ameaça de desaparecimento vivenciado pela maioria das línguas indígenas no Brasil, os dicionários multimídia surgem como uma ferramenta para auxiliar na revitalização de uma língua e podem ser úteis tanto no processo de transmissão das línguas quanto em seus processos de documentação e descrição.

PALAVRAS-CHAVE
Línguas indígenas. Lexicografia. Revitalização de línguas.

ABSTRACT

This article discusses the importance of multimedia dictionaries as multifunctional and multipurpose products developed as a social technology aimed at the learning of endangered languages. Multimedia dictionaries for Indigenous languages represent a reusable, freely available, and easily accessible methodology that can be used by communities on a large scale without the need for internet access. The use of technology such as smartphones, tablets, or computers, though still incipient, is already a reality in many Indigenous communities, which favors the implementation of dictionaries in digital formats whose lexical entries can be illustrated with multimedia resources. The article first presents the concept, methodology, and technology involved in creating these dictionaries and describes the macro- and microstructures adopted in them, which are determined according to the needs of speakers or potential speakers of the language. It also discusses the role of multimedia dictionaries in language planning within the context of revitalization practices for minoritized and endangered languages, based on the experience of developing dictionaries for eight Indigenous languages at the Linguistics Division of the Museu Paraense Emílio Goeldi (MPEG) in collaboration with the University of New Mexico (UNM, USA): Kanoé, Sakurabiat, Oro Win, Djeoromitxi, Puruborá, Makurap, Wayoro and Salamãi. In the context of vulnerability and the threat of extinction faced by most Indigenous languages in Brazil, multimedia dictionaries emerge as a tool to support language revitalization and can be useful both in the transmission of languages and in their documentation and description.

KEY-WORDS
Indigenous languages. Lexicography. Language revitalization.

RESUMO PARA NÃO ESPECIALISTAS

O cenário de desaparecimento experimentado pela ampla maioria das línguas indígenas faladas no Brasil vem mobilizando cada vez mais esforços, seja das comunidades de fala destas línguas ou da comunidade científica, em torno de práticas de revitalização linguística, isto é, ações voltadas para manter uma língua em uso. Este tipo de prática, muitas vezes, requer ferramentas que auxiliem no processo para que essas ações sejam mais eficazes e alcancem os objetivos desejados. O presente trabalho se propõe a discutir uma dessas ferramentas, os dicionários multimídias de línguas indígenas, uma ferramenta digital que pode ser utilizada nas ações voltadas para a manutenção do uso de uma língua indígena. O trabalho apresenta a metodologia utilizada para a elaboração de dicionários multimídias digitais e discute a experiência de oito dicionários de línguas indígenas em elaboração no setor de linguística do Museu Paraense Emílio Goeldi em colaboração com a Universidade do Novo México, focando nos usos e nas características estruturais de cada um. Além disso, o trabalho também discute a inserção desta ferramenta no planejamento linguístico das comunidades e o papel desempenhado pelos

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.13693

dicionários multimídias digitais como ferramentas de suporte à revitalização das línguas indígenas.

Introdução

Este artigo aborda a importância dos dicionários multimídia como um produto multifuncional e multiuso, desenvolvido como uma tecnologia social reaplicável, de uso livre e fácil acesso, que pode ser usada pelas comunidades em larga escala e sem necessidade de acesso à internet. Tecnologia social[1] é definida como um conjunto de saberes, metodologias, técnicas, conhecimentos científicos e tradicionais integrados para buscar solução para um ou vários problemas sociais. No caso do tema deste artigo, reverter ou reduzir as perdas linguísticas. O trabalho discute as propostas de elaboração de dicionários multimídia de línguas minorizadas dentro de uma perspectiva voltada para a aprendizagem e revitalização das línguas, tendo como enfoque o planejamento linguístico necessário para a definição do escopo dos dicionários que são desenvolvidos no contexto de uma metodologia de base comunitária (*community-based methodology*) em que as comunidades de falantes estejam envolvidas no processo. De modo geral, a metodologia para a elaboração dos dicionários multimídia para línguas indígenas (DMLI) abordada neste trabalho consiste em compilar dados linguísticos e arquivos de mídia em um formato que possibilita um fácil compartilhamento e acesso (Brito; Birchall; Galucio, 2025).

Os Dicionários Multimídias para Línguas Indígenas aliam documentação de línguas indígenas e tecnologia digital de comunicação e informação, e podem ser usados em aplicativos móveis (celulares, tablets, computadores) sem necessidade de internet. A tecnologia é de utilização gratuita, com aplicação de softwares de acesso livre aberto, e pode ser utilizada pelas comunidades, nas escolas e em ambientes fora do espaço escolar. Essa Tecnologia Social envolve desenvolvimento de metodologias reaplicáveis e de produtos digitais, visando contribuir para a documentação de línguas indígenas bem como fomentar sua aprendizagem e revitalização. O uso de tecnologia como smartphones, tablets ou computadores, ainda que incipiente, é uma realidade em muitas comunidades indígenas, o que favorece a implementação de dicionários em formatos digitais, cujas entradas lexicais podem ser ilustradas com recursos multimídias (além de informações conceituais, definições, transcrição, campos semânticos, exemplos de uso etc.).

[1] A Rede de Tecnologia Social (RTS) define tecnologias sociais como "produtos, técnicas ou metodologias reaplicáveis, desenvolvidas na interação com a comunidade e que representem efetivas soluções de transformação social".

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.13693

O trabalho apresenta a metodologia para elaboração dessa modalidade de dicionários, desenvolvida no setor de linguística do Museu Paraense Emílio Goeldi (MPEG) em colaboração com a Universidade de Novo México (UNM, EUA) e discute o papel desse tipo de dicionário no planejamento linguístico associado a práticas de revitalização de línguas indígenas minorizadas e ameaçadas, a partir da experiência de elaboração de dicionários multimídia para oito línguas indígenas, a saber: Kanoé, Sakurabiat, Oro Win, Djeoromitxi, Puruborá, Makurap, Wayoro e Salamãi.

O restante do artigo está organizado da seguinte forma. Na seção 1**,** apresentamos a metodologia e a tecnologia envolvidas na elaboração dos dicionários multimídia, concebidos como uma tecnologia social para apoio à aprendizagem e revitalização de línguas. A seção 2 apresenta o contexto sociolinguístico das línguas Kanoé, Sakurabiat, Oro Win, Puruborá, Djeoromitxi, Wayoro, Makurap e Salamãi e a metodologia, rotinas de trabalhos e tecnologia utilizadas no processo de criação dos dicionários multimídias para cada uma dessas línguas, que foram as primeiras aplicações da metodologia de desenvolvimento dos DMLI. Na seção 3 abordamos o papel dos dicionários multimídias enquanto ferramenta didática no processo de transmissão linguística. Apresentamos as principais contribuições do artigo na seção 4.

1. Metodologia: Dicionários multimídia para línguas indígenas (DMLI) - uma tecnologia social com foco na aprendizagem e revitalização de línguas

Os dicionários são um dos principais instrumentos para descrever uma língua, junto com as coleções de textos e as gramáticas. A sua elaboração parte do registro do conhecimento lexical dos falantes, ou seja, das palavras e expressões que compõem o vocabulário de uma comunidade. Um dicionário pode ter muitos usos: servir como apoio no ensino, como fonte para estudos sobre o léxico ou como referência para a própria comunidade linguística. Seu alcance também pode variar: pode ser mais restrito, abordando apenas um tema específico, como nomes de plantas ou objetos da cultura material, ou mais amplo, reunindo informações de diferentes áreas do léxico. Para uma discussão mais detalhada sobre lexicografia e a produção de dicionários de línguas indígenas, ver Birchall (2023).

Dicionários multimídia são aqueles que incorporam recursos como clipes de áudio e vídeo em suas entradas, oferecendo exemplos de pronúncia das palavras e ilustrações de seu uso em frases mais complexas. Um dos papéis principais de um dicionário é servir como referência para a comunidade linguística. Quando essas comunidades contam com poucos falantes ativos capazes de modelar a pronúncia e o uso das palavras, os recursos multimídia

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.13693

tornam-se especialmente valiosos. Eles permitem registrar e transmitir de forma mais fiel aspectos orais da língua, como pronúncia e contexto de uso, funcionando não apenas como ferramenta de referência, mas também como instrumento de preservação e revitalização linguística. Esses formatos digitais apresentam vantagens em relação aos formatos impressos, como a distribuição e o fácil acesso dos *app*s entre usuários de smartphones ou tablets, redução de custos e flexibilidade na organização da macroestrutura. Outra vantagem dessa metodologia é o versionamento, através do qual a equipe pode lançar atualizações graduais à comunidade de usuários do dicionário. Assim o dicionário pode passar por várias etapas de revisão e a comunidade pode ter acesso ao material de forma imediata.

A proposta da tecnologia social de Dicionários Multimídia para Línguas Indígenas discutida neste artigo é integrar a elaboração desses dicionários ao próprio fluxo de trabalho de documentação linguística. Nesse sentido, busca-se aproveitar de materiais comumente presentes em acervos de línguas, como bancos de dados lexicais e transcrições temporalmente alinhadas de recursos multimídia, de modo a facilitar a construção do dicionário e potencializar o aproveitamento do trabalho já realizado.

Para apoiar e guiar o processo de elaboração de DMLI, estabelecemos algumas práticas gerais, como a estrutura de bancos de dados lexicais e convenções de nomear arquivos multimídia, e desenvolvemos alguns recursos computacionais, como o aplicativo CSV2RMD. Essa metodologia faz parte da documentação do aplicativo CSV2RMD (Brito; Birchall; Galucio 2025) e será detalhada a seguir[2].

O aplicativo CSV2RMD permite gerar dicionários digitais a partir de arquivos CSV (*Comma Separated Values*) padronizados e preenchidos com dados lexicais e um conjunto organizado de arquivos de mídia, que podem ser convertidos tanto para o formato HTML (*HyperText Markup Language*) quanto para PDF (*Portable Document Format*). Um dicionário em formato HTML permite uma organização mais dinâmica do conteúdo, com vantagens como a possibilidade de reestruturação dinâmica e a inclusão de mídias como áudio e vídeo, que não poderiam ser utilizadas diretamente em outros formatos mais tradicionais, além do encapsulamento desse formato em APK para dispositivos Android. Já o documento em PDF, embora não possibilite a manipulação dinâmica de sua estrutura ou a adição de mídias em áudio

[2] Um outro modelo de elaboração de dicionários multimídia para línguas indígenas no Brasil é aquele aplicado nos dicionários disponibilizados no Portal Japiim (https://japiim.museudoindio.gov.br/), desenvolvido no âmbito do Projeto de Documentação de Línguas Indígenas (ProDoclin) do Museu dos Povos Indígenas em parceria com a Unesco (http://prodoclin.museudoindio.gov.br/).

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.13693

ou vídeo, permite criar dicionários em um formato mais similar aos tradicionais, que podem ser impressos e distribuídos onde a utilização de dispositivos eletrônicos não está disponível.

As ferramentas computacionais utilizadas no processo de elaboração desses recursos multimídia são os softwares de transcrição e anotação linguísticas (ELAN, 2025), o uso de *scripts* em Python para a automatização na extração de dados segmentados (texto e áudio) de arquivos ELAN (.eaf) para planilhas CSV e posterior conversão para HTML, e ainda uma distribuição do LaTeX chamada MiKTeX, necessária para a produção de arquivos PDF. O fluxo de trabalho utilizado abrange desde a organização dos dados até a exportação dos produtos e pode ser sumarizado em algumas etapas: **preparação de dados e metadados, configuração e verificação, e criação de produtos**.

A **preparação de dados e metadados** envolve a organização dos materiais necessários para a criação dos produtos, seguindo modelos previamente definidos. Essa etapa inclui organização das mídias em pastas padronizadas e preenchimento do banco de dados lexical em formato CSV, que pode ser elaborado manualmente ou gerado a partir de informações transcritas em arquivos no formato *.eaf* utilizando um script CSV2RMD desenvolvido para essa finalidade. Cada entrada no banco de dados lexicais deve conter, no mínimo, três campos obrigatoriamente preenchidos: Item Lexical, Tradução/Significado e Campo Semântico. Além disso, podem ser incluídos campos opcionais, tais como: Imagem, Arquivo Sonoro, Transcrição Fonêmica, Transcrição Fonética, Classe Gramatical, Descrição, Arquivo Sonoro do Exemplo, Transcrição do Exemplo, Tradução do Exemplo, Arquivo de Vídeo, Subcampo Semântico e Itens Relacionados. A macro e microestrutura de cada dicionário são definidas conforme a necessidade dos falantes ou potenciais falantes de cada língua.

Com os dados devidamente organizados, o usuário pode realizar a ***configuração e verificação***, na qual, a partir do primeiro uso do programa principal, pode validar o preenchimento da tabela CSV, buscando erros e corrigindo-os, ou ainda adicionar arquivos de configuração que permitem a customização e o preenchimento dos documentos multimídia gerados. Através desses arquivos de configuração, é possível, por exemplo, customizar a ordem alfabética ou adicionar uma introdução textual ao documento.

Uma vez validado e configurado na etapa anterior, o ambiente de trabalho pode ser utilizado para a etapa final de ***criação de produtos***, que permite gerar (através do programa principal) documentos HTML ou PDF (figura 1). Esses documentos podem, então, ser revisados e analisados para divulgação ou ainda ajustados, caso sejam necessárias modificações, reiniciando assim um novo ciclo de atividades.

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.13693

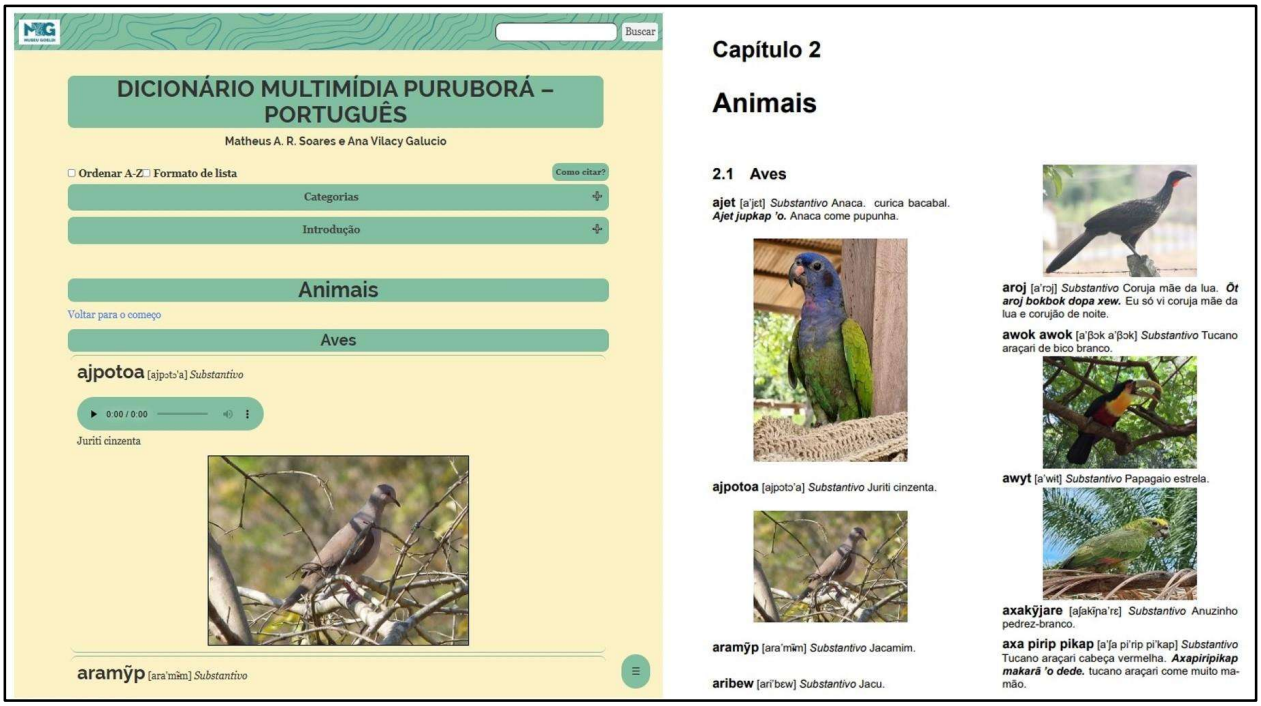


Figura 1: Amostras da versão HTML online (esquerda) e PDF (direita) do Dicionário Puruborá (Soares et al. 2024).

Os produtos HTML gerados podem conter arquivos de mídia embutidos ou vinculados externamente. A principal vantagem da mídia embutida é permitir o acesso ao conteúdo sem necessidade de conexão à internet. Por outro lado, isso aumenta significativamente o tamanho do arquivo final no dispositivo do usuário. A produção de aplicativos Android, com o HTML do dicionário integrado, ainda não está totalmente consolidada, pois depende do uso de um script externo online para sua geração. Atualmente, esses aplicativos podem ser compartilhados de maneira informal, mas não podem ser publicados em lojas oficiais de aplicativos. Essa etapa do fluxo de trabalho segue em desenvolvimento, com o objetivo de automatizar a criação de arquivos APK que atendam aos requisitos técnicos necessários para serem disponibilizados em plataformas como a Google Play Store a partir da versão HTML do dicionário.

Todas as ferramentas e práticas aqui descritas foram utilizadas na produção de dicionários e essas atividades, por sua vez, possibilitaram diversas melhorias nas próprias práticas, permitindo hoje a distribuição em livre acesso de manuais e ferramentas que viabilizam a construção simplificada de DLMI's, com baixa necessidade de conhecimento técnico. As ferramentas estão disponíveis e podem ser acessadas na plataforma GitHub que contém um repositório destinado ao registro e desenvolvimento do sistema, permitindo a

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.13693

colaboração entre os interessados na utilização e melhoria desses processos (Brito; Birchall; Galucio, 2025)[3].

A aplicação dos procedimentos metodológicos de elaboração dos DMLI aqui descritos permite ao usuário gerar tabelas de dados e recortar áudios correspondentes a itens lexicais e seus exemplos de uso a partir de anotações padronizadas no ELAN, realizar ajustes em imagens e áudios, bem como a produção dos arquivos em PDF e HTML. A aplicação dessa metodologia será ilustrada na próxima seção, através de oito exemplos de projetos de dicionários multimídia, desenvolvidos no âmbito da cooperação entre o Museu Paraense Emílio Goeldi e a Universidade de Novo México, para o desenvolvimento dos DMLI.

2. Dicionários multimídia: aplicação da metodologia

Nesta seção, apresentamos oito estudos de caso de aplicação da metodologia de desenvolvimento dos dicionários multimídia (Brito; Birchall; Galucio, 2025), a saber: os dicionários das línguas Kanoé, Sakurabiat, Oro Win, Puruborá, Djeoromitxi, Makurap, Wayoro e Salamãi. Inicialmente, descreve-se brevemente a caracterização do povo e a situação sociolinguística de cada uma das línguas indígenas discutidas no artigo e, em seguida, apresenta-se o processo de produção e a caracterização do dicionário de cada língua, destacando a abordagem colaborativa com as respectivas comunidades de fala.

A documentação das línguas é realizada por pesquisadores indígenas e não-indígenas em parceria com falantes nativos das línguas específicas. Alguns DMLIs produzidos seguindo a metodologia e usando a tecnologia descritos neste artigo aproveitaram da documentação linguística disponível no Acervo de Línguas Indígenas do Museu Paraense Emílio Goeldi (ALIM), como o dicionário multimídia Kanoé - Português. O ALIM é uma coleção de valor inestimável, caracterizando-se como um acervo multimídia que reúne uma vasta documentação sobre a diversidade linguística da Amazônia. Este acervo é um pilar fundamental para as pesquisas e a preservação do patrimônio linguístico e cultural dos povos indígenas da região. Atualmente o acervo possui em seus registros informações de 80 línguas, compondo um volume total de 2,74 TB de informações. Para os outros dicionários, foi possível coletar dados com membros falantes das respectivas comunidades em trabalho colaborativo, seja em visitas nas aldeias, seja através de visita dos falantes ao laboratório de Linguística do Museu Paraense

[3] Ver em https://github.com/MPEG-Linguistica/CSV2RMD. O conteúdo do repositório inclui scripts que podem ser baixados em formato para Windows (.exe) ou em Python, um manual de utilização com o passo a passo, disponível em formato HTML e PDF.

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.13693

Emílio Goeldi, conforme será apresentado a seguir. Os cinco primeiros dicionários (Kanoé, Sakurabiat, Oro Win, Puruborá e Salamãi) já estão concluídos e em uso pelas respectivas comunidades e podem ser acessados através do portal de Dicionários Multimídia do Museu Paraense Emílio Goeldi[4]. Já os dicionários das línguas Djeoromitxi, Makurap e Wayoro estão em fase de finalização e estão previstos para serem disponibilizados no portal no primeiro semestre de 2026.

2.1 Dicionário Kanoé

Os Kanoé reportam que seu território tradicional se localizava na região dos rios Corumbiara e Tanaru, na região sul do Estado de Rondônia. Atualmente, encontram-se dispersos em vários municípios e terras indígenas, onde convivem com outras etnias, sendo que a maioria da população se encontra distribuída em quatro terras indígenas localizadas no sul do estado de Rondônia: Terra Indígena Rio Guaporé, Terra Indígena Rio Branco, Terra Indígena Sagarana e Terra Indígena Omerê. Desde a época dos primeiros contatos com a população não-indígena, o povo Kanoé tem sido marcado por histórias trágicas, envolvendo doenças e agressões, que provocaram uma significativa redução populacional e afetaram também a vitalidade da língua tradicional do povo. A língua Kanoé (kxo), classificada como uma língua isolada (Rodrigues, 1986), é hoje pouquíssimo falada, tendo deixado de ser transmitida há cerca de duas gerações. Somente na Terra Indígena Rio Omerê ainda há um pequeno grupo de pessoas da etnia, composto por 3 adultos, que são falantes fluentes da língua e a usam na comunicação diária. Após o falecimento de dois anciãos Kanoé habitantes da Terra Indígena Rio Guaporé, há notícias de somente mais um falante fluente nesta região. Nas outras duas terras indígenas citadas acima, não há informação da presença de falantes fluentes da língua. Entretanto, a comunidade Kanoé está interessada em resgatar sua língua tradicional e vem desenvolvendo estratégias voltadas para sua revitalização (Aquino; Kanoe, 2024).

Nesse contexto é que foi desenvolvido o Dicionário Multimídia Kanoé – Português (Marcon; Galucio; Brito, 2023), elaborado a partir de uma demanda específica da comunidade Kanoé da Terra Indígena Rio Guaporé, através do cacique José Augusto Kanoé. Sua demanda foi por uma ferramenta que auxiliasse no processo de aprendizagem da língua, mas que possibilitasse escutar a pronúncia das palavras, uma vez que, na sua percepção, materiais somente escritos não satisfaziam a necessidade da comunidade. Este foi o primeiro dicionário produzido utilizando a metodologia dos DMLIs. O dicionário foi elaborado com base no acervo

[4] Ver em https://dicionarios.museu-goeldi.br/

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.13693

da língua, disponível no Arquivo de Línguas Indígenas do Museu Goeldi – ALIM (versão 1.0 foi produzida em 2020). Foram transcritas gravações de itens lexicais e frases realizadas pelo linguista Denny Moore (MCTI/Museu Goeldi), com D. Teresa Piraguê, na década de 1990, e por Ana Vilacy Galucio (MCTI/Museu Goeldi), com o Sr. Francisco Kanoé, em 2008, na Terra Indígena Rio Guaporé. O dicionário contém 457 entradas lexicais, sendo que uma pequena porcentagem das entradas conta também com frases ilustrativas do contexto de uso da palavra. A macroestrutura do dicionário está organizada por campos semânticos, divididos em 13 categorias: animais, corpo, cores, eventos, localizações, manufaturas, natureza, números, parentescos, pessoas, plantas, propriedades e outros. A microestrutura do dicionário Kanoé é composta pela entrada lexical na ortografia Kanoé, seguida da tradução em português, transcrição fonética em Kanoé, classe gramatical, aúdio com a pronúncia da palavra em Kanoé, uma imagem representativa, e sempre que possível uma frase com a palavra em contexto e o respectivo áudio da frase. Porém, devido à limitação dos dados disponíveis, a porcentagem de entradas contendo exemplos ilustrativos de uso é relativamente pequena.

2.2 Dicionário Sakurabiat

A língua Sakurabiat (skf) pertence ao ramo Tupari, da família linguística Tupi, e é a língua tradicional do povo Sakurabiat[5], que vive na região do estado de Rondônia, majoritariamente na Terra Indígena Rio Mequens. Segundo dados de campo, levantados por Ana Vilacy Galucio, com o falecimento do cacique Geraldinho Guaratira Saquirabiar, em março de 2025, há somente 08 falantes fluentes e cerca de uma dezena de falantes parciais da língua. O número reduzido de falantes e o fato de não estar sendo transmitida para as novas gerações, colocam a língua Sakurabiat entre as línguas indígenas brasileiras severamente ameaçadas. Apesar de a língua não estar sendo transmitida para as novas gerações há pelo menos três décadas (Galucio, 2021), há interesse entre os jovens adultos de fortalecer a língua e retomar sua aprendizagem (Guaratira; Costa, 2020; Galucio, 2020).

Neste contexto de interesse pela aprendizagem da língua, vinculado ao ensino da mesma no sistema escolar, os professores indígenas e outros membros da comunidade solicitaram consultoria e auxílio dos linguistas para a elaboração de material didático-pedagógico para alfabetização e ensino-aprendizagem da língua dentro e fora do espaço escolar

[5] Na literatura e em documentos pessoais (RG, CPF) e oficiais o nome da etnia também ocorre registrado como Sakyrabiar e Saquirabiar. A língua era referida na literatura antiga como Mekens ou Mequens, mas nas últimas décadas o povo adotou a autodenominação Sakurabiat e mais recentemente tem também se referido às variedades da língua ainda faladas na T.I. Rio Mequens, ou seja, as variedades Guaratira e Sakurabiat.

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.13693

(Costa, 2020a). Foi inicialmente produzido um livro de apoio à alfabetização e aprendizagem da língua (Costa, 2020b) e iniciada a elaboração do dicionário multimídia Sakurabiat-Português, cuja primeira versão foi elaborada em 2022 e após revisão e validação pelos falantes e outros membros da comunidade, foi publicada e disponibilizada para uso pela comunidade em junho de 2024.

O Dicionário Multimídia Sakurabiat–Português foi elaborado a partir de corpus da coleção digital Sakurabiat, que está hospedada no ALIM/MPEG[6], e de registros complementares de áudios para as entradas lexicais e frases ilustrativas de uso dos verbetes, com falantes da língua, durante etapas de pesquisa de campo na Terra Indígena Rio Mequens e no Museu Paraense Emílio Goeldi. A versão atual do dicionário Sakurabiat-Português (Galucio et al., 2024) apresenta 646 itens lexicais, organizados em uma macroestrutura composta por 11 campos semânticos: animais, subdivididos em aves, insetos, mamíferos, peixes, répteis, anfíbios, e outros animais; plantas; natureza; corpo; eventos; manufaturas; pessoa; propriedades; e outros. O dicionário apresenta, ainda, um índice alfabético. A microestrutura de cada verbete inclui, conforme a disponibilidade de dados, o item lexical, sua transcrição fonética, a indicação da classe gramatical, tradução correspondente para o português, bem como exemplos de uso contextualizados em formato de áudio e texto. Além disso, os verbetes contam com imagens ilustrativas e gravações em áudio, tanto da palavra isolada quanto de frases ilustrativas de uso do verbete em questão.

2.3 Dicionário Oro Win

O povo Oro Win é da região da cabeceira do rio Pacaás Novos, situada no município de Guajará-Mirim, no estado de Rondônia. A população étnica atual é de aproximadamente 150 pessoas, a maioria das quais reside na Terra Indígena Uru-Eu-Wau-Wau. Apenas seis pessoas falam a língua com fluência nativa, e menos de vinte possuem alguma competência produtiva (Birchall, 2018). Todos os falantes nativos têm mais de 60 anos, enquanto os demais falantes geralmente têm entre 30 e 60 anos. A língua ainda é usada como meio de comunicação entre os anciãos e, ocasionalmente, entre estes e os outros falantes. Nos últimos anos, a comunidade Oro Win tem implementado um programa de ensino da língua tradicional nas escolas locais (Birchall; Oro Win Cabixi, 2021).

[6] A coleção da língua Sakurabiat no Acervo de Línguas Indígenas do Museu Goeldi (ALIM) vem sendo criada ao longo dos anos, a partir de projetos de documentação e descrição da língua coordenados por Ana Vilacy Galucio, desde a década de 1990. Esse trabalho resultou em uma coleção com cerca de 150 horas de gravação audiovisual, distribuídas em diferentes gêneros discursivos e elicitações de léxico e gramática.

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.13693

Em 2019, teve início o projeto do Dicionário Oro Win, financiado pelo programa *Documenting Endangered Languages* (DEL) dos Estados Unidos, por meio de uma bolsa concedida ao linguista Joshua Birchall. O objetivo principal do projeto foi capacitar pesquisadores Oro Win em métodos de documentação linguística, de modo que eles pudessem registrar, em formato audiovisual, o conhecimento léxico dos falantes da língua e elaborar um dicionário multimídia em parceria com o linguista[7]. Dois pesquisadores Oro Win, Silvânia Oro Eo' Cabixi e Luciano Oro Win, foram indicados pela comunidade, após demonstrarem proficiência na compreensão e escrita da língua. Eles receberam treinamento em técnicas de condução de entrevistas linguísticas para elicitação de dados lexicais, bem como no uso do programa ELAN para transcrição. Com a chegada da pandemia de COVID-19, no início de 2020, a presença do linguista na aldeia tornou-se inviável até o lançamento da primeira versão do dicionário em 2022. Nesse período, os pesquisadores indígenas assumiram a continuidade da documentação linguística dentro da própria comunidade, demonstrando como uma metodologia de pesquisa de base comunitária pode gerar resultados não apenas relevantes do ponto de vista científico, mas também fortalecedores do protagonismo indígena na preservação e transmissão de sua própria língua e cultura.

A versão mais recente do Dicionário Oro Win (3.0) foi entregue à comunidade em junho de 2025, reunindo 858 entradas lexicográficas (Birchall; Oro Eo' Cabixi; Oro Win, 2025). Cada entrada apresenta a transcrição ortográfica, fonêmica e fonética do item lexical, além de informações como classe gramatical, observações culturais, gravação de áudio com a pronúncia e um exemplo contextualizado de uso, também acompanhado de áudio. O dicionário reúne exemplos extraídos de narrativas do acervo Oro Win disponíveis no ALIM, bem como de trabalhos de campo voltados à elicitação de itens lexicais, realizados pelos autores do dicionário. Após uma revisão conduzida com falantes da língua em julho de 2025, está em preparação uma versão final para publicação formal impressa por uma editora científica de acesso aberto, bem como para a atualização da próxima versão do dicionário multimídia online.

2.4 Dicionário Puruborá

Puruborá é a língua tradicional do povo Puruborá, originário da região do atual estado de Rondônia, onde até hoje vive a maioria da sua população. Uma parte do povo vive na aldeia

[7] A coleção de documentação Oro Win no ALIM vem sendo desenvolvida em parceria com a comunidade linguística desde 2009. Atualmente essa coleção Oro Win inclui cerca de 58 horas de produtos audiovisuais, distribuídos entre narrativas, diálogos e elicitações. Os dados coletados entre 2019 e 2022 como parte do projeto DEL também foram depositados no *Archive of Indigenous Languages of Latin America* (AILLA) na Universidade do Texas.

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.13693

Aperoi, localizada às margens da rodovia BR-429, nas proximidades do rio Manuel Correia, município de Seringueiras (RO) e outra parte vive em diversos outros municípios do mesmo estado. Devido ao longo e violento processo de exploração vivenciado pelos Puruborá, a língua do povo sofreu uma profunda quebra de transmissão intergeracional e interrupção em seu uso e não foi falada durante muitos anos, chegando quase a desaparecer (Galucio, 2005). Com o falecimento de três anciãos, entre 2020 e 2023, há somente um ancião que ainda lembra algumas palavras da língua e um novo falante, o professor Mário Oliveira Neto que vem aprendendo a língua nas últimas duas décadas a partir dos produtos oriundos da documentação linguística, e alunos-aprendentes que têm estudado a língua no contexto de fortalecimento e revitalização cultural desenvolvido pelos Puruborá (Oliveira Neto, 2020; Galucio, 2020). A língua é o único membro conhecido do ramo Puruborá, família Tupi, e compõe um subagrupamento com a língua Karo denominado Ramarama-Puruborá (Galucio; Gabas Júnior, 2002; Rodrigues; Cabral, 2012).

O dicionário multimídia bilíngue Puruborá - Português foi elaborado a partir de corpus da coleção digital Puruborá, que está hospedada no ALIM/MPEG[8], e de registros complementares de áudios para as entradas lexicais e frases ilustrativas de uso dos verbetes. A versão atual do dicionário possui um total de 727 itens lexicais divididos em uma macroestrutura composta por 11 campos semânticos: animais (com as devidas subdivisões entre aves, insetos, mamíferos, peixes, répteis e anfíbios), corpo, eventos, manufaturas, natureza, números, pessoas, plantas, propriedades, saudações e outros. Todos os 727 itens lexicais possuem tradução para a língua portuguesa e os respectivos áudios dos falantes pronunciando a palavra. Além disso, 424 itens lexicais possuem imagem e 507 possuem exemplos de uso. A microestrutura do dicionário da língua Puruborá é composta por item lexical, transcrição fonética, classe gramatical, tradução, áudio do item lexical, imagem, exemplo de uso, tradução do exemplo e áudio do exemplo. Tal microestrutura é preenchida conforme a disponibilidade de informações, havendo, por conta disto, itens lexicais que não possuem todos estes campos preenchidos. Após validação e verificação pelos usuários, em 2023, a versão mais recente do dicionário Puruborá (Soares et al., 2024) foi disponibilizada para uso da comunidade Puruborá, em 2024 e está sendo utilizada pelos membros da comunidade, tanto no contexto escolar na aldeia Aperoi, nas aulas de aprendizagem da língua

[8] Resultado de um projeto de documentação da língua Puruborá, coordenado por Ana Vilacy Galucio, no período de 2001 a 2007, a Coleção Puruborá no Acervo do Museu Goeldi possui cerca de 91 horas de gravação audiovisual, distribuídas entre elicitações, relatos pessoais e tradicionais do povo.

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.13693

e cultura tradicional Puruborá, quanto nas aulas online de revitalização da língua e nos espaços externos à escola (cf. Mário Oliveira Neto, comunicação pessoal, 19 de setembro de 2025).

2.5 Dicionário Djeoromitxi

Djeoromitxi foi classificada como uma língua pertencente ao agrupamento Macro-Jê (Ribeiro; van der Voort, 2010). Os Djeoromitxi (230 indivíduos, segundo registro do SIASI/SESAI do ano de 2020) estão distribuídos em dois territórios indígenas no sul de Rondônia, a T. I. Rio Guaporé e a T. I. Rio Branco (onde há parte da população, mas não há registro de falantes). O contato dos Djeoromitxi com a sociedade não-indígena intensificou-se no início do século XX, durante o ciclo da borracha, provocando grandes mudanças culturais. A expedição de Emil-Heinrich Snethlage nos anos 1930 marcou os primeiros registros sistemáticos sobre a língua e cultura do grupo (ver Snethlage, 1939; Ribeiro; van der Voort, 2010). A língua Djeoromitxi é considerada "severamente ameaçada" pela UNESCO devido à falta de transmissão da língua entre gerações. O censo do INDL/IPHAN realizado pela equipe do Museu Goeldi mostrou que 35 falam fluentemente, além disso, identificou que entre 50 e 60 pessoas possuem fluência passiva, grupo que poderia ser foco, além da retomada da transmissão intergeracional, para uma intervenção no processo de revitalização da língua (Moore, 2018).

Em 2022, teve início o projeto de documentação da língua Djeoromitxi, financiado pelo Endangered Languages Documentation Programme/ELDP (Rocha, 2022), originando uma ampla coleção audiovisual, incluindo 5 horas de textos anotados (transcritos e traduzidos) passíveis de integrar o dicionário multimídia com exemplos de uso contextual. O dicionário Djeoromitxi e uma gramática pedagógica (ambos em fase de revisão) são produtos resultantes do trabalho de documentação linguística[9].

O dicionário multimídia Djeoromitxi foi elaborado por Juliana Solano e Ivan Rocha, com a comunidade Djeoromitxi, a partir de corpus da coleção digital Djeoromitxi, que está hospedada no ALIM/MPEG e no ELAR ARCHIVE. A versão atual do dicionário encontra-se em fase de revisão e validação pela comunidade e deve estar disponível para uso em 2026. A metodologia adotada para a elaboração do dicionário pode ser descrita em quatro eixos principais: (i) coleta de dados em campo com falantes da língua na Terra Indígena Rio Guaporé

[9] A coleção Djeoromitxi, no Acervo de Línguas Indígenas do Museu Goeldi (ALIM), vem sendo constituída desde a década de 1990 pelos linguistas Denny Moore, Hein van der Voort, Nádia Pires e Thiago Castro. O projeto de documentação Djeoromitxi (Rocha, 2022) ampliou esse acervo, com aproximadamente 170 horas de produtos audiovisuais, organizados em diferentes gêneros discursivos e elicitações de léxico e gramática.

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.13693

e uso de listas de palavras do Acervo ALIM/MPEG; (ii) segmentação e transcrição de mídias de áudio (iii) estruturação da base lexicográfica em formato digital por campos semânticos (fauna, flora, parentesco, advérbios, objetos, partes do corpo, fenômenos climáticos, entre outros); e (iv) processamento técnico para a geração do produto final.

Cada entrada lexical do dicionário multimídia foi organizada de forma a integrar informações linguísticas, pedagógicas e multimodais. Como exemplo, apresenta-se o item lexical *hibzi* 'chicha'. A entrada contempla: a forma ortográfica padronizada da palavra em Djeoromitxí, conforme a proposta de Pires (1992); a transcrição fonética [ˈhi.bzi], que registra sua realização sonora; a classificação gramatical como "nome", por designar uma entidade concreta; e a tradução para o português, possibilitando a identificação do equivalente mais próximo. Além disso, inclui uma frase exemplificadora, que demonstra o uso do vocábulo em contexto e evidencia sua função sintática e semântica na oração. O material também é composto por um arquivo sonoro, gravado por falantes Djeoromitxi, e por uma ilustração, a fim de reforçar a associação visual e apoiar o processo de aprendizagem. Por fim, a palavra é inserida no campo semântico "bebidas", categoria responsável por organizar o vocábulo dentro de um grupo temático específico.

Esse processo metodológico permitiu não apenas a construção de um dicionário com mais de 800 entradas lexicais com informações multimídia, mas também a integração entre documentação linguística, desenvolvimento tecnológico e devolução do conhecimento à comunidade, bem como o envolvimento de membros da comunidade no processo.

2.6 Dicionário Wayoro

Tradicionalmente localizados nas cabeceiras dos rios ˀUgu ngop (Rio Colorado) e Tõremõ ãpi (Rio Terebinto), ambos no Estado de Rondônia (Brasil), os Wajuru (ou Ngwayoro ou Wayoro 'povo da Pedra') tiveram os primeiros contatos com os não indígenas na primeira metade do século XX. Desse contato se desencadeou uma catastrófica epidemia de sarampo, escravização por empresas seringalistas e o deslocamento forçado de seus territórios tradicionais. Atualmente, o povo Wajuru habita duas regiões do estado de Rondônia: Porto Rolim de Moura do Guaporé (município de Alta Floresta d'Oeste) e a Terra Indígena Rio Guaporé (município de Guajará-Mirim). A população atual é de cerca de 340 indivíduos. A língua Wayoro não é mais falada no dia a dia. As crianças Wajuru se comunicam em português. Há apenas uma senhora idosa (77 anos de idade) que pode ser considerada falante fluente de Wayoro, além de cerca de vinte semi-falantes (Nogueira, 2019).

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.13693

Em 2023, iniciou-se um projeto de documentação sistemática da língua e da cultura Wayoro, através do projeto *"Documentation of the severely endangered languages Makurap and Wayoro (Brazil): material and non-material traditional culture, and its associated knowledge"*, apoiado pelo Endangered Languages Documentation Programme (Wajuru et al., 2025; Nogueira, Wajuru; Wajuru, 2023)[10]. Concomitantemente à documentação linguística, são realizadas descrição e análises linguísticas que embasam a elaboração de materiais didáticos e paradidáticos. Documentação e arquivamento, descrição e análise linguística, bem como a elaboração de materiais de ensino de língua constituem um fluxo de trabalho cíclico retroalimentar, colaborativo, permanente e dinâmico (Nogueira; Massaka; Wajuru, 2023). Atendendo à solicitação da comunidade Wajuru, foram elaborados dois materiais didáticos de ensino da língua Wayoro (não publicados) e um vocabulário Wayoro-Português (Nogueira; Wajuru; Sousa, 2021). Atualmente, além do dicionário multimídia Wajuru - Português, uma equipe de pesquisadores indígenas está desenvolvendo, sob a consultoria da Dra. Antônia Fernanda de Souza Nogueira, um livro de ensino da língua Wayoro, com o apoio do Museu dos Povos Indígenas/UNESCO.

O dicionário multimídia Wayoro-Português conta, atualmente, com 643 entradas lexicais, organizadas em 11 campos semânticos, a saber: animais, plantas, natureza, corpo, cores, pessoas, comidas, itens do dia a dia, estados e propriedades, atividades e eventos, e outros. Cada entrada lexical apresenta: a palavra escrita na língua indígena; classe de palavra; áudio da palavra na língua Wayoro; equivalente em português; imagem correspondente; sentença exemplo de uso na ortografia Wayoro; equivalente em português da sentença exemplo; áudio da sentença exemplo de uso. Entretanto, por questões de disponibilidade, alguns itens podem não conter todos os recursos listados. A versão atual do dicionário Wayoro (versão 1.0) ainda está em processo de finalização, com previsão de ser disponibilizada para validação e teste pela comunidade de fala no primeiro semestre de 2026.

2.7 Dicionário Makurap:

O povo Makurap é formado por uma população estimada em cerca de 500 pessoas, distribuídas nas terras indígenas Rio Guaporé e Rio Branco, no estado de Rondônia (Brasil).

[10] Os primeiros registros de áudio da língua Wayoro na Coleção Wayoro do ALIM foram realizados nos anos finais da década de 1990, por pesquisadores do Museu Paraense Emílio Goeldi. Em 2008, Antônia Fernanda de Souza Nogueira iniciou uma parceria com a etnia Wajuru da Terra Indígena Rio Guaporé, realizando documentação linguística, especialmente em forma de elicitação, de maneira intermitente. Atualmente, com o resultado do projeto de documentação, o Acervo Wayoro possui cerca de 148 horas de gravação audiovisual, distribuídos entre elicitações, como a gravação de listas de palavras e frases do dia a dia da comunidade, relatos pessoais e tradicionais do povo.

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.13693

Classificada como parte do ramo Tupari, da família Tupi, a língua Makurap assumiu, no passado, o status de língua franca na região do Guaporé (Maldi, 1991). Atualmente, porém, encontra-se em situação de ameaça devido à combinação de fatores históricos violentos, consequentes do contato com não-indígenas. Por consequência, o número de falantes foi reduzido para aproximadamente 24 a 30 pessoas e a transmissão intergeracional da língua foi interrompida (Wajuru et al., 2025)[11].

Entretanto, um forte interesse da comunidade de fala em aprender a língua Makurap foi despertado por meio de um amplo projeto colaborativo de documentação da língua e cultura Makurap, o que tem motivado a comunidade de fala a iniciar esforços autônomos de revitalização da língua e da cultura Makurap (Wajuru et al., 2025). Entre as iniciativas propostas pelo povo Makurap está a realização da "Festa do Peixe *Potkap Xako*", organizada pela *Awanda* Associação dos Povos Makurap do Rio Guaporé. A festividade, realizada em setembro de 2025, teve como tema a revitalização da cultura e como objetivos valorizar a língua e a cultura Makurap, além de incentivar atividades culturais entre o povo. Iniciativas como essa demonstram o empenho dos Makurap em participar de projetos que possam contribuir positivamente para a revitalização de sua língua de tradição.

O dicionário multimídia Makurap-Português contém, atualmente, 1.075 entradas lexicais. A macroestrutura do dicionário consiste na organização dos lexemas em 14 categorias semânticas. Já a microestrutura inclui, principalmente: lema, transcrição fonética, frases contextuais (em Makurap, com tradução para o português), imagem ilustrativa e áudios de pronúncia da palavra e das frases e, geralmente, apenas um equivalente de tradução para o português. Entretanto, segundo Rocha e Galucio (no prelo), foi necessário, para alguns casos específicos, realizar um estudo lexical demandado por questões específicas da língua Makurap (casos de variação, inserção de palavras gramaticais, lematização de verbos, homonímia, polissemia e classificação semântica preliminar dos itens lexicais de fauna e flora) para as quais determinadas soluções analíticas influenciaram a composição dos verbetes e a disposição das entradas lexicais, a partir das estratégias de organização de dicionários possíveis na metodologia proposta por Brito, Birchall e Galucio (2025). Uma das questões consideradas neste estudo foi a inclusão de palavras gramaticais (posposições, partículas, pronomes e

[11] A coleção Makurap, como parte do Acervo de Línguas Indígenas do Museu Goeldi (ALIM), vem sendo constituída desde a década de 1990, através de ações esporádicas de documentação pelos linguistas Denny Moore e Ana Vilacy Galucio. Somando ao material pré-existente, desenvolvemos um novo projeto de documentação, financiado pelo Endangered Languages Documentation Programme/ELDP (Galucio; Nogueira; Costa, 2023), originando uma coleção de aproximadamente 160 horas de gravação audiovisual, organizados em diferentes gêneros discursivos e elicitações de léxico e gramática.

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.13693

conectivos) no dicionário, em que se buscou decidir como incluir nos verbetes as informações necessárias para que o usuário possa compreender a função gramatical dessas palavras na estrutura da língua utilizando os recursos opcionais da planilha CSV, a saber: descrição e itens relacionados. Esse estudo foi fundamental para estabelecer critérios específicos sobre como apresentar, no dicionário multimídia, as relações e nuances lexicais e gramaticais identificadas no trabalho lexicográfico de forma consistente e padronizada (Rocha; Galucio, no prelo). A versão atual do dicionário Makurap (versão 1.0) ainda está em processo de finalização, com previsão de ser disponibilizada para validação e teste pela comunidade de fala no primeiro semestre de 2026.

2.8 Dicionário Salamãi

Os Salamãi tradicionalmente viviam na região entre os rios Pimenta Bueno e Barão de Melgaço, no estado de Rondônia. Atualmente, membros da comunidade vivem em vários municípios do estado, como Porto Velho, Guajará-Mirim e Chupinguaia. Tradicionalmente falam a língua Salamãi (mnd), pertencente ao ramo Mondé da família Tupí, embora em 2021 houvesse registro de apenas duas mulheres idosas que ainda se lembravam de palavras do idioma. Tratava-se de uma sociedade multiétnica, com fortes laços culturais e sociais com os Aikanã e os Kwaza, incluindo práticas como as *chichada*s e as trocas matrimoniais. Na literatura, são frequentemente referidos como "Mondé", denominação que na realidade correspondia ao nome de um cacique, mas que acabou sendo usada também para designar a família linguística à qual o Salamãi pertence (Moore, 2005). Entre as décadas de 1930 e 1950, os Salamãi mantiveram contato com seringueiros ao longo do rio Pimenta Bueno e, nos anos 1960, muitos foram forçados a trabalhar na construção da BR-364 (Cuiabá–Porto Velho). Parte do grupo permaneceu em Porto Velho, onde contribuiu para a fundação da Reserva Indígena Cassupá/Salamãi.

A documentação linguística que serve de base ao Dicionário Salamãi insere-se no contexto do projeto *A Composição Etnolinguística do Sudoeste de Rondônia*, desenvolvido no âmbito do programa internacional DoBeS (*Dokumentation Bedrohter Sprachen*), financiado pela Fundação Volkswagen. Esse projeto produziu um registro multimídia abrangente das línguas e culturas indígenas do sudeste de Rondônia — especialmente dos povos Aikanã, Kwaza e seus vizinhos imediatos, entre eles os Salamãi, incluindo narrativas tradicionais, histórias pessoais, diálogos, cantos e dados linguísticos. As gravações utilizadas especificamente na elaboração do Dicionário Salamãi foram realizadas por Joshua Birchall, em 2016 e 2019 com Dona Peridalva Wanzerip, uma das últimas pessoas que falam a língua

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.13693

Salamãi. As gravações consistem principalmente em sessões de levantamento de palavras isoladas e formas verbais flexionadas, além de registros de músicas e histórias pessoais. Todo o material processado até o momento encontra-se depositado no Acervo de Línguas Indígenas do Museu Paraense Emílio Goeldi (ALIM) e no acervo DoBeS dentro de *The Language Archive*, garantindo sua preservação e acesso para futuras pesquisas. Durante o trabalho de campo, Birchall restabeleceu contato com familiares de Dona Peridalva que vivem na região do Guaporé, que expressaram o desejo de que os materiais coletados fossem utilizados para apoiar o estudo de sua língua.

O Dicionário Salamãi foi elaborado a partir dos dados coletados por Joshua Birchall e trabalhados posteriormente por Bruno Pinto Silva, doutorando em Linguística pela Universidade do Novo México (Pinto Silva; Birchall, 2024). A elaboração do dicionário seguiu essas etapas, com o auxílio do software ELAN: (1) segmentação das palavras em Salamãi, (2) transcrição ortográfica, (3) tradução para o português conforme as interpretações de Dona Peridalva, (4) extração e seleção dos trechos de áudio correspondentes, (5) compilação dos sons transcritos com símbolos do Alfabeto Fonético Internacional (IPA), e (6) identificação de padrões fonológicos a partir do corpus.

O dicionário reúne 114 itens lexicais provenientes de cinco arquivos de áudio, abrangendo os principais campos semânticos: plantas e alimentos (25 itens), animais (73) e parentesco (16). A macroestrutura organiza os verbetes segundo esses domínios temáticos, enquanto a microestrutura de cada entrada inclui o lema na ortografia do Salamãi, a transcrição fonética em IPA, a classe gramatical, a tradução em português e o arquivo de áudio correspondente. Sempre que possível, foram incluídas informações sobre variação morfológica e, nos casos de verbos, formas flexionadas. Convenções ortográficas foram estabelecidas de modo a manter relativa proximidade com o português, favorecendo a legibilidade e o uso por falantes bilíngues. A definição do sistema gráfico e a análise fonológica preliminar basearam-se em referências como Hanke (1950), Galucio et al. (2015) e Felzke e Moore (2019). Embora o corpus ainda não permita uma análise detalhada da tonalidade, foram marcadas vogais longas pela duplicação da letra correspondente. O resultado é um produto lexicográfico bilíngue, multimodal e acessível, que alia rigor linguístico e relevância comunitária.

3. Discussão: Dicionários multimídia e transmissão linguística

Esta seção discute o papel dos dicionários multimídias enquanto ferramenta didática no processo de transmissão linguística, incluindo casos específicos de retomada e/ou revitalização

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.13693

linguística. Ressalta-se sua importância como recurso pedagógico, sem presumir que a simples elaboração de um dicionário para uma língua ameaçada seja suficiente para documentar, preservar ou revitalizar tal idioma.

Dentre as cerca de 150 línguas indígenas faladas atualmente no Brasil, cerca de 40% a 50% está ameaçada e a outra parte em situação de vulnerabilidade. Nos últimos anos, muitos povos indígenas cujas línguas encontram-se em situação de vulnerabilidade, com poucos falantes, espaços de uso restritos e ruptura na transmissão intergeracional, iniciaram ou planejam iniciar movimentos autônomos de retomada, ou revitalização de suas línguas tradicionais. Nesse contexto, uma questão essencial é o planejamento linguístico, que inclui a avaliação e a descrição do status (situação), do corpus (material existente, documentado e descrito) e da aquisição (modo de transmissão linguística e produtos disponíveis) (Sallabank, 2012; McCarty, 2018; Amaral, 2020). Por exemplo, a preparação de material de apoio didático utilizado no processo de transmissão da língua é exemplo de planejamento da aquisição.

A revitalização linguística exige ações integradas de *status*, corpus e aquisição, orientadas pelo protagonismo das comunidades afetadas e apoiadas por políticas sensíveis ao contexto linguístico, histórico e sociocultural. Nestas etapas do planejamento, os dicionários multimídia surgem como uma ferramenta para auxiliar na revitalização de uma língua e podem ser tanto úteis no processo de transmissão da língua quanto documental e descritivo, através da integração entre o trabalho dos linguistas e das comunidades indígenas para a documentação das línguas e produção de materiais acessíveis para quem usa essas línguas.

Nesse cenário, em 2019, o Museu Paraense Emílio Goeldi, que atua há décadas na documentação e estudo das línguas dos povos indígenas no Brasil, iniciou o desenvolvimento dos dicionários multimídia para línguas indígenas (DMLI), respondendo à demanda dos grupos indígenas por materiais para ajudar na manutenção e aprendizagem de suas línguas de tradição. Os dicionários multimídia apresentados neste artigo foram planejados como ferramentas para contribuir com ações comunitárias de revitalização de línguas, aplicação de técnicas linguísticas e o uso de tecnologia, demonstrando a cooperação respeitosa, sem imposição, entre as comunidades atendidas e os linguistas envolvidos no processo e foram desenvolvidos para auxiliar nas iniciativas de revitalização linguística. Inicialmente, a tecnologia dos dicionários multimídia foi desenvolvida para atender a uma demanda de povos indígenas de Rondônia, em uma colaboração entre linguistas e as comunidades indígenas para a formulação e elaboração de produtos para ajudar no fortalecimento e revitalização de suas línguas tradicionais.

Conforme mencionado nas seções anteriores, os Dicionários Multimídias combinam documentação de línguas indígenas e tecnologia digital de comunicação e informação. A

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.13693

Tecnologia Social envolve desenvolvimento de metodologias reaplicáveis e de produtos digitais, visando contribuir para a documentação de línguas indígenas e fomentar sua aprendizagem e revitalização, através da integração entre o trabalho dos linguistas e das comunidades indígenas para a documentação das línguas e produção de materiais acessíveis para quem usa essas línguas. O uso de tecnologia como smartphones, tablets ou computadores favorece a implementação de dicionários em formatos digitais, cujas entradas lexicais podem ser ilustradas com recursos multimídias.

Assim, a prática da dicionarização e o formato multimídia, que pode ser usado no celular, tablet ou computador, *online* ou *offline*, contribui para aumentar a motivação dos falantes e potenciais falantes, inclusive crianças e jovens, uma vez que o uso do dicionário possibilita a ampliação do vocabulário da língua e a familiaridade com os sons e a pronúncia das palavras. Os oito dicionários multimídia apresentados neste artigo estão sendo elaborados com esse alinhamento, ou seja, os membros das comunidades envolvidos em todo o processo de produção dos dicionários, especialmente no que diz respeito ao planejamento do corpus, à validação do produto e ao uso da ferramenta nas ações em suas comunidades.

Essa tecnologia social é projetada para apoiar ativamente os processos de transmissão e aquisição linguística na comunidade. Esta é a função concreta dos dicionários. Além disso, os dicionários multimídia fomentam a promoção ou melhoria, de certa forma, da autoestima da comunidade e da percepção positiva dos falantes ou futuros aprendizes com relação a suas línguas. Ver suas línguas dentro dessa tecnologia, conforme temos experienciado, muitas vezes causa emoção nos usuários, especialmente quando a tecnologia permite trazer de volta as vozes de falantes que já não estão mais no meio da comunidade. Isto graças aos minuciosos trabalhos de documentação e à preservação de longa-duração dos acervos digitais, como o Acervo de Línguas Indígenas do Museu Goeldi (ALIM), que permitiu o resgate do material sonoro (também audiovisual) que foi utilizado para o desenvolvimento dos dicionários das línguas Kanoé, Puruborá, Djeoromitxi, Oro Win, Sakurabiat e Salamãi.

A tecnologia e o trabalho incansável de documentar as línguas também permitiram colocar nesse tipo de mídia as vozes de falantes que ainda estão vivos na comunidade. Isto também gera outro tipo de motivação, trazendo à tona uma discussão muito importante no processo de revitalização, que é aproveitar a presença e disponibilidade dos falantes como uma fonte para a transmissão linguística, enquanto eles estão aqui. Concretamente, os usuários entendem que as ferramentas multimídia são fontes importantes de consultas para o aprendizado, mas que quando é possível consultar um falante e obter o contexto de uso daquele

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.13693

conhecimento linguístico direto de um avô, avó, tia etc., através da interação social, o aprendizado se torna mais eficaz.

Por outro lado, é fundamental reconhecer que a simples existência de um dicionário, por mais tecnologicamente avançado que seja, não garante por si só a preservação ou revitalização de uma língua. Para que cumpra seu papel de forma eficaz, seu desenvolvimento deve ser intencionalmente alinhado às três dimensões do planejamento linguístico: corpus, status e aquisição. No contexto das línguas abordadas, os dicionários nascem de um minucioso planejamento de **corpus**, que envolve a coleta, documentação e estruturação do léxico e contextos de uso. Simultaneamente, eles visam fortalecer o **status** da língua ao apresentá-la em um formato moderno, digital e acessível (via aplicativos em smartphones), o que aumenta sua visibilidade e prestígio, especialmente entre os jovens. O objetivo final, no entanto, é o planejamento da **aquisição**: servir como um instrumento prático que facilite o aprendizado e a transmissão intergeracional.

Para alcançar esse objetivo, a concepção do dicionário deve ir além da base de dados do linguista. É preciso um planejamento pedagógico que considere o público-alvo. Isso implica em decisões cruciais, como a seleção de campos lexicais relevantes para os aprendizes, a formulação de exemplos de uso claros e contextualmente ricos, e a incorporação estratégica de recursos multimídia (áudio, imagens, vídeos). Esses elementos transformam o dicionário de uma lista de palavras e textos naturalísticos em uma experiência de imersão que apoia a aprendizagem da pronúncia, do significado e do uso culturalmente apropriado da língua.

Contudo, a eficácia pedagógica do dicionário depende diretamente da precisão de seu conteúdo. O processo de coleta lexical, especialmente em domínios especializados como fauna e flora, revela complexas relações entre o signo linguístico (a palavra na língua indígena), seu significado (a tradução) e o referente (a espécie ou objeto no mundo). A experiência com o dicionário Wayoro-Português ilustra bem esses desafios. Por exemplo:

- A entrada *nangoro* foi traduzida como ‘camaleão’, mas a imagem associada era a de um lagarto comum na região, gerando uma imprecisão biológica, visto que camaleões não são nativos das Américas e que os animais popularmente referidos como camaleões são espécies de lagartos.
- A expressão *‘ugu djori* (lit. ‘jacamim da água’) foi aplicada pela falante a duas aves distintas (andorinha-do-mar e pavãozinho-do-pará), sendo que a primeira é uma ave marinha improvável na região de Rondônia.
- A palavra *akui* foi inicialmente identificada a partir da imagem de uma barata-do-mato, mas a mesma colaboradora, em um contexto comunitário, corrigiu a tradução

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.13693

para uma ‘espécie de grilo’, evidenciando como mesmo o estímulo visual pode levar a identificações equivocadas.

Esses exemplos demonstram que uma entrada imprecisa, seja na tradução ou na associação com uma imagem, pode levar a uma aprendizagem equivocada, comprometendo o objetivo central da transmissão linguística. Para evitar tais riscos, a documentação lexical exige uma metodologia rigorosa e, idealmente, uma colaboração interdisciplinar. Como sugerem Fleck (2007) e Berto (2010), a parceria com biólogos, o uso de guias de campo específicos da fauna e flora locais, estímulos auditivos, como o canto de aves e anfíbios, são importantes e até essenciais para garantir a correta identificação dos referentes.

Quatro dos dicionários apresentados neste artigo (Kanoé, Sakurabiat, Oro Win e Puruborá) já estão sendo utilizados pelas respectivas comunidades de fala, como ferramenta de auxílio para aprendizagem e revitalização das línguas. O dicionário da língua Kanoé foi distribuído para uso pelos membros da comunidade no ambiente familiar e residencial como uma ferramenta de consulta e aprendizagem, tanto na Terra Indígena Rio Guaporé quanto na Terra Indígena Rio Branco. No caso do dicionário Sakurabiat, além do uso pelos membros da comunidade no ambiente familiar e residencial, também está sendo usado como ferramenta de aprendizagem nas duas escolas atuais da Terra Indígena Rio Mequens. O dicionário multimídia da língua Puruborá vem sendo incluído gradativamente nas atividades relacionadas à revitalização da língua. Inicialmente, o dicionário havia sido distribuído para os membros da comunidade, os quais vinham utilizando este produto no ambiente familiar e residencial como uma ferramenta de consulta. Mais recentemente, está ocorrendo um planejamento para um uso mais sistemático desta ferramenta no ambiente escolar como suporte ao processo de ensino-aprendizado da língua e alinhada a outros materiais didáticos da língua. O uso do dicionário Oro Win já foi implementado na sala de aula de duas escolas da comunidade Oro Win desde 2022. Também foram produzidos quatro cartazes a partir do dicionário para auxiliar os professores indígenas no ensino da língua, abrangendo temas como o uso do alfabeto e os nomes de vários tipos de plantas e animais.

Conforme mencionado anteriormente, os resultados diretos da aplicação e do uso das ferramentas e procedimentos metodológicos para o desenvolvimento dos dicionários multimídia para línguas indígenas (DMLI) descritos neste artigo podem ser conferidos no site de dicionários do Museu Paraense Emílio Goeldi (https://dicionarios.museu-goeldi.br/), criado para divulgar os produtos gerados a partir da aplicação da metodologia junto às comunidades estudadas e o público em geral. O site contém atualmente seis dicionários publicados. Nele, é

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.13693

possível visualizar o conteúdo online ou baixar versões de cada um desses produtos em HTML, PDF ou para Android (APK), quando disponíveis.

4. Considerações finais

No cenário atual no Brasil de forte ameaça à diversidade linguística e considerando a motivação e o interesse de vários povos indígenas pela retomada, revitalização e fortalecimento das suas respectivas línguas tradicionais, os dicionários multimídia se consolidam como uma poderosa ponte entre a documentação linguística e a revitalização de base comunitária.

Como apresentado nas seções anteriores, a tecnologia e metodologia para elaboração dos DMLI apresentadas neste artigo utiliza softwares de uso aberto, é reaplicável, de uso livre e fácil acesso e pode ser usada pelas comunidades em larga escala e sem necessidade de acesso à internet, podendo também ser aplicada a qualquer língua. Sua utilização pelas comunidades cujos dicionários estão em uso demonstram que os dicionários multimídia podem funcionar como ferramenta didática no processo de transmissão linguística e que, nesse caso, não podem funcionar como meros repositórios lexicais, mas podem ir além e cumprir um papel de ferramenta (junto com outros materiais de apoio didático) nas ações de revitalização. Nesse sentido é que a abordagem metodológica e a aplicação da tecnologia na produção dos dicionários são concebidas como uma tecnologia social e como uma ferramenta pedagógica dinâmica que pode ser atualizada constantemente devido à abordagem colaborativa entre comunidade, linguistas e profissionais da área de tecnologia.

Considerando que intervenções de fortalecimento ou retomada devem partir de um desejo e objetivo das comunidades, que podem ser apoiadas por agentes institucionais quanto ao uso de tecnologia e metodologias sistematizadas e replicáveis, com o objetivo de facilitar a transmissão, o sucesso dos dicionários multimídia não reside apenas na tecnologia empregada, mas na integração cuidadosa entre o planejamento de corpus, status e aquisição, e em um compromisso contínuo com a precisão e a relevância cultural do conteúdo, garantindo que ele seja uma ferramenta confiável e eficaz para os atuais e futuros falantes.

Agradecimentos

Agradecemos a todos as comunidades indígenas envolvidas nos projetos de produção de dicionários multimídia pela parceria e colaboração. Agradecemos as seguintes instituições que apoiaram as pesquisas que possibilitaram o desenvolvimento do presente artigo: *National Endowment for the Humanities (EUA),* programa *Documenting Endangered Languages* (projeto FN-266285-19 *Documentation and Dictionary of Oro Win (orw)*, coordenado por Joshua Birchall). *Endangered Languages Documentation Programme* (projetos *MDP0435 Documentation of the severely endangered languages Makurap and Wayoro (Brazil): material*

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.13693

and non-material traditional culture, and its associated knowledge, coordenado por Ana Vilacy Galucio e Antonia Fernanda Nogueira; e projeto *IPF0386 Documentation and description of the Djeoromitxí language from Amazonia, Brazil*, coordenado por Ivan Rocha da Silva). Conselho Nacional de Pesquisa e Desenvolvimento/CNPq (bolsa de produtividade em pesquisa (PQ)- 304766/2022-4 para Ana Vilacy Galucio Processo; bolsa de capacitação institucional (PCI) no MPEG/MCTI-314025/2025-1 para Saulo Brito; bolsa de Iniciação Científica no CNPq/MPEG para Douglas Rodrigues Junior, Israel Marcon, Flávio Pinheiro, Juliana Solano, Matheus Soares e Victor Rocha). Fundação Nacional dos Povos Indígenas/FUNAI: apoio logístico durante as pesquisas de campo e autorizações de ingresso em área indígena, processos 08620.002095/2009-57, 08620.010349/2022-88 e 08620.009160/2022-42.

Pesquisa com seres humanos

As pesquisas reportadas neste artigo contaram com o consentimento informado de todos os participantes. Os projetos de pesquisa específicos que incluíram parte das pesquisas reportadas neste artigo foram aprovados pela Comissão Nacional de Ética em Pesquisa, sob os números CAAE 63101422.1.0000.0173 e CAAE 62625822.8.0000.0173.

REFERÊNCIAS

AMARAL, Luiz. Estratégias para a revitalização de línguas ameaçadas e a realidade brasileira. **Cadernos de Linguística**, *[S. l.]*, v. 1, n. 3, p. 01–44, 2020. DOI: 10.25189/2675-4916.2020.v1.n3.id251. Disponível em: https://cadernos.abralin.org/index.php/cadernos/article/view/251. Acesso em: 8 out. 2025.

AQUINO, Letícia de Souza; KANOE, M. **Compartilhando dados e experiências de Revitalização Linguística com a comunidade Kanoé do Guaporé**. 2024. (Apresentação de Trabalho/Comunicação).

BERTO, Flávia de Freitas. **Nomes de aves em Jurunas: um estudo lexicológico.** Monografia (Bacharel em Letras) - Universidade Estadual Paulista "Júlio de Mesquita Filho", Campus de Araraquara. 2010. Disponível em: https://repositorio.unesp.br/server/api/core/bitstreams/0344b232-4211-407d-a7e6-df145259567e/content. Acesso em: 30 sep. 2025.

BIRCHALL, Joshua. Lexicografía y elaboración de diccionarios. *In:* P. ARANDIA (org.) **Introducción a la Lingüística**: Curso para investigadores de lenguas indígenas de Bolivia, Cochabamba: Página y Signos, 2023. p. 141-159.

BIRCHALL, Joshua. **Oro Win**. Inventário Nacional da Diversidade Linguística. Relatório técnico apresentado ao IPHAN/Museu Paraense Emílio Goeldi, 2018. (manuscrito).

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.13693

BIRCHALL, Joshua; ORO EO' CABIXI, Silvania; ORO WIN, Luciano. **Dicionário Oro Win.** Versão 3.0. 2025. Disponível em: https://dicionarios.museu-goeldi.br/dic/oro_win/. Acessado em: 07/10/2025

BIRCHALL, Joshua; ORO WIN CABIXI, Olivia. Passos na retomada da língua do povo Oro Win. *In:* TONDINELI, Patrícia Goulart. (ed.), **(Re)vitalizar línguas minorizadas e/ou ameaçadas:** teorias, metodologias, pesquisas e experiências, Porto Velho: EDUFRO, 2021. p. 69-82.

BRITO, Saulo; BIRCHALL, Joshua; GALUCIO, Ana Vilacy. **Csv2rmd**: Um programa python para produzir dicionário multimídia com Markdown. Versão 1.0. 2025. DOI: https://doi.org/10.5281/zenodo.14894549.

COSTA, Carla Nascimento da**. Material Didático Sakurabiat**. (Desenvolvimento de material didático ou instrucional - Material Didático). MPEG: Belém. 2020a.

COSTA, Carla Nascimento da. **Proposta de Material Didático para a Língua Sakurabiat**. 160 f. Dissertação (Mestrado) – Programa de Pós-graduação em Letras, Universidade Federal do Pará, Belém, 2020b.

ELAN (Versão 7.0) [Software de computador]. Nijmegen: Instituto Max Planck de Psicolinguística, **The Language Archive**, 2025. Disponível em:<https://archive.mpi.nl/tla/elan>. Acesso em: 07 out. 2025

FELZKE, Lediane Fani; MOORE, Denny. Kinship terminology of the groups of the Mondé language family. **Boletim do Museu Paraense Emílio Goeldi,** Ciências Humanas, v. 14, n. 1, p. 15-32, 2019. DOI: 10.1590/1981.81222019000100003.

FLECK, David. Field linguistics meets biology: how to obtain scientific designations for plant and animal names. **STUF - Language Typology and Universals**, Berlin, vol. 60, no. 1, 2007, pp. 81-91. https://doi.org/10.1524/stuf.2007.60.1.81

GALUCIO, Ana Vilacy. Documentação e Revitalização Linguística: uma interseção possível, necessária e desejável. *In*: Patrícia Goulart Tondineli. (Org). **(Re)vitalizar línguas minorizadas e/ou ameaçadas: teorias, metodologias, pesquisas e experiências**. 1ed. Porto Velho: EDUFRO, 2021, v.1, p. 20 - 43.

GALUCIO, Ana Vilacy. Heritage languages, linguistic obsolescence and motivations for learning in the context of indigenous school education: reflections, doubts, and challenges. **Cadernos de Linguística**, *[S. l.]*, v. 1, n. 3, p. 01–20, 2020. DOI: 10.25189/2675-4916.2020.v1.n3.id235. Disponível em: https://cadernos.abralin.org/index.php/cadernos/article/view/235. Acesso em: 30 sep. 2025.

GALUCIO, Ana Vilacy. Puruborá: notas etnográficas e linguísticas recentes. Boletim do Museu Paraense Emílio Goeldi. Ciências Humanas, Belém, v. 1, n.2, p. 159-192, 2005. 2005

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.13693

GALUCIO, Ana Vilacy; GABAS JÚNIOR, Nilson. Evidências de agrupamento genético Karo-Puruborá, tronco Tupi [Evidence for the genealogical grouping Karo-Puruborá, Tupi stock]. Paper presented at the XVII **Encontro Nacional da Associação Nacional de Pós-graduação de Letras e Linguística**, Gramado, Brazil, 2002.

GALUCIO, Ana Vilacy; MEIRA, Sérgio; BIRCHALL, Joshua; MOORE, Denny; GABAS JÚNIOR, Nilson; DRUDE, Sebastian; STORTO, Luciana; PICANÇO, Gessiane; RODRIGUES, Carmen Reis. Genealogical relations and lexical distances within the Tupian linguistic family. **Boletim do Museu Paraense Emílio Goeldi, Ciências Humanas,** v. 10, 2015. Disponível em: https://doi.org/10.1590/1981-81222015000200004. Acesso em: 30 sep. 2025

GALUCIO, Ana Vilacy; NOGUEIRA, Antonia Fernanda; COSTA, Carla Daniele da. **Makurap: Documentation of Language and Culture**, 2023. Endangered Languages Archive. Handle: http://hdl.handle.net/2196/73b9o01s-1c78-2246-5187-0fm978700a80. Acesso em: 07 out. 2025.

GALUCIO, Ana Vilacy; RODRIGUES JUNIOR, Douglas; BRITO, Saulo (com Manoel Sakyrabiar, Olimpio Sakyrabiar, Rosalina Guaratira, Severino Guaratira e Matheus A. R. Soares). 2024. **Dicionário multimídia Sakurabiat-Português.** Versão 3.0. Belém: Museu Paraense Emílio Goeldi. Disponível em http://dicionarios.museu-goeldi.br/dic/sakurabiat/. Acesso em 28 set. 2025.

GUARATIRA, Silvana da Silva Cunha; COSTA, Carla Nascimento da. Experiences of Sakurabiat language and culture rescue. **Cadernos de Linguística**, *[S. l.]*, v. 1, n. 3, p. 01–14, 2020. DOI: 10.25189/2675-4916.2020.v1.n3.id249. Disponível em: https://cadernos.abralin.org/index.php/cadernos/article/view/249

HANKE, Wanda. Breves notas sobre os índios Mondé e o seu idioma. **Dusenia** (Curitiba), vol. I, n. 4, 1950, pp. 215-228.

MALDI, Denise. O Complexo Cultural do Marico: Sociedades Indígenas dos Rios Branco, Colorado e Mequens, Afluentes do Médio Guaporé. **Boletim do Museu Paraense Emílio Goeldi**, Série Antropologia, vol. 7(2), pp. 209-269, 1991.

MARCON, Israel.; GALUCIO, Ana Vilacy; Brito, Saulo. 2023. **Dicionário multimídia Kanoe-Português.** Versão 2.0. Belém: Museu Paraense Emílio Goeldi. Disponível em http://dicionarios.museu-goeldi.br/dic/kanoe/. Acesso em 09 out. 2025.

McCARTY, Teresa. Community-based language planning: Perspectives from indigenous language revitalization. In : HINTON, L.; HUSS, L. M.; ROCHE, G. (Ed.). **The Routledge handbook of language revitalization**. New York and London: Routledge, 2018, p. 22 –35.

MOORE, Denny. Classificação interna da família lingüística Mondé. **Estudos Lingüísticos** XXXIV. p. 515-520, 2005. Disponível em: https://etnolinguistica.wdfiles.com/local--files/journal:estudos/moore_2005_monde.pdf. Acesso em: 28 sep. 2025.

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.13693

MOORE, Denny. **Djeoromitxi**. Inventário Nacional da Diversidade Linguística. Relatório técnico apresentado ao IPHAN/Museu Paraense Emílio Goeld, 2018. (manuscrito).

NOGUEIRA, Antônia Fernanda de Souza. **Predicação na língua Wayoro (Tupi)**: propriedades de finitude. 189 f. Tese (Doutorado em Semiótica e Lingüística Geral) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2019. Disponível em: https://www.teses.usp.br/teses/disponiveis/8/8139/tde-14082019-101006/en.php. Acesso em: 25 sep. 2025.

NOGUEIRA, Antonia Fernanda de Souza; MASSAKA, Gilson Wajuru; WAJURU, Adão. 'We Didn't Wear Skirts': Extralinguistic Factors in the Development of Wayoro Pedagogical Materials. In: VALIJÄRVI, Riitta-Liisa; KAHN, Lily (Eds.). **Teaching and Learning Resources for Endangered Languages**. Leiden: Brill, 2023. p. 190 - 202.

NOGUEIRA, Antonia Fernanda de Souza; WAJURU, Comunidade; SOUSA, Clenilson M. Apresentação da primeira versão do vocabulário Wayoro-Português In: TONDINELI, Patrícia Goulart (Org.). **(Re)vitalizar línguas minorizadas e/ou ameaçadas: teorias, metodologias, pesquisas e experiências**. Porto Velho: EDUFRO, 2021. p. 83 - 108.

OLIVEIRA NETO, Mario. Revitalization and teaching of the Puruborá language. **Cadernos de Linguística**, *[S. l.]*, v. 1, n. 3, p. 01–16, 2020. DOI: 10.25189/2675-4916.2020.v1.n3.id246. Disponível em: https://cadernos.abralin.org/index.php/cadernos/article/view/246. Acesso em: 30 sep. 2025.

PINTO SILVA, Bruno; BIRCHALL, Joshua. **Dicionário multimídia Salamãi-Português**. Versão 1.0. 2024. Disponível em http://dicionarios.museu-goeldi.br/dic/salamai/. Acesso em: 07 out 2025.

PIRES, Nádia Nascimento. **Estudo da gramática da língua Jeoromitxi (Jabutí)**: Aspectos sintáticos das cláusulas matrizes. 157 f. Dissertação (Mestrado em Linguística), Universidade Estadual de Campinas, Campinas, 1992.

RIBEIRO, Eduardo Rivail; VAN DER VOORT, Hein. Nimuendajú was right: The inclusion of the Jabutí language family in the Macro-Jê stock. **International Journal of American Linguistics**, v. 76, n. 4, p. 517-570, 2010. DOI: https://doi.org/10.1086/658056.

ROCHA, Ivan. Documentation and description of the Djeoromitxí language from Amazonia, Brazil. **Endangered Languages Archive**, 2022. Disponível em: http://hdl.handle.net/2196/2ef60f9c-796c-4b6f-c50c-417343074e7g. Acesso em: 02 out. 2025.

ROCHA, Victor Siqueira; GALUCIO, Ana Vilacy. Dicionário Multimídia para a língua Makurap: Uma Ferramenta para Aprendizagem da Língua. **Letras Escreve**. No prelo.

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.13693

RODRIGUES, Aryon Dall'Igna. CABRAL, Ana Suelly Arruda. Câmara. Tupían. In: CAMPBELL, Lyle; Grondona, Verónica. (Org.). The Indigenous Languages of South America: a comprehensive guide. Berlin: Mouton de Gruyter, 2012, p. 495-573.

RODRIGUES, Aryon Dall'Igna. Línguas Brasileiras: para o conhecimento das línguas indígenas. São Paulo: Edições Loyola, 1986.

SALLABANK, Julia. From language documentation to language planning: Not necessarily a direct route. In : SEIFART, F. et al. (ed.). Potentials of language documentation: Methods, analyses, and utilization. Honolulu: University of Hawaii Press, **Language Documentation & Conservation**, v. 3. p. 118 – 125, 2012.

SNETHLAGE, Emil Heinrich. Musikinstrumente der Indianer des Guaporégebietes. **Baessler-Archiv**, Beiträge zur Völkerkunde, vol. 10. Berlin: Dietrich Reimer and Andrews & Steiner. 1939.

SOARES, Matheus Augusto Ribeiro.; GALUCIO, Ana Vilacy; OLIVEIRA NETO, Mario (com Paulo Aporete Filho, José EvangelistaPuruborá e Saulo Brito). 2024. **Dicionário multimídia Puruborá-Português.** Versão 3.0. Belém: Museu Paraense Emílio Goeldi. Disponível em http://dicionarios.museu-goeldi.br/dic/purubora/. Acessado em 28/09/2025.

SOARES, Matheus Augusto Ribeiro; RODRIGUES JUNIOR, Douglas da Costa; GALUCIO, Ana Vilacy. Dictionary Making as a Tool for the Revitalization of Endangered Languages: Two Case Studies. **Cadernos de Linguística**, *[S. l.]*, v. 5, n. 1, p. e704, 2024. DOI: 10.25189/2675-4916.2024.v5.n1.id704. Disponível em: https://cadernos.abralin.org/index.php/cadernos/article/view/704. Acesso em: 1 oct. 2025

WAJURU, Jacqueline; MACURAP, Jociclei; NOGUEIRA, Antonia Fernanda; GALUCIO, Ana Vilacy; COSTA, Carla Nascimento. Engaging Wajuru /Wayoro and Makurap Communities in Collaborative Documentation: Recording, Learning, and Communicating. **Dutkansearvvi dieđalaš áigečála,** v. 09, p. 54-74, 2025.

Este preprint foi submetido sob as seguintes condições:

- Os autores declaram que os necessários Termos de Consentimento Livre e Esclarecido de participantes ou pacientes na pesquisa foram obtidos e estão descritos no manuscrito, quando aplicável.
- Os autores declaram que a elaboração do manuscrito seguiu as normas éticas de comunicação científica.
- Os autores declaram que estão cientes que são os únicos responsáveis pelo conteúdo do preprint e que o depósito no SciELO Preprints não significa nenhum compromisso de parte do SciELO, exceto sua preservação e disseminação.
- Os autores declaram que os dados, aplicativos e outros conteúdos subjacentes ao manuscrito estão referenciados.
- O manuscrito depositado está no formato PDF.
- Os autores declaram que a pesquisa que deu origem ao manuscrito seguiu as boas práticas éticas e que as necessárias aprovações de comitês de ética de pesquisa, quando aplicável, estão descritas no manuscrito.
- Os autores declaram que uma vez que um manuscrito é postado no servidor SciELO Preprints, o mesmo só poderá ser retirado mediante pedido à Secretaria Editorial do SciELO Preprints, que afixará um aviso de retratação no seu lugar.
- Os autores concordam que o manuscrito aprovado será disponibilizado sob licença Creative Commons CC-BY.
- O autor submissor declara que as contribuições de todos os autores e declaração de conflito de interesses estão incluídas de maneira explícita e em seções específicas do manuscrito.
- Os autores declaram que o manuscrito não foi depositado e/ou disponibilizado previamente em outro servidor de preprints ou publicado em um periódico.
- Caso o manuscrito esteja em processo de avaliação ou sendo preparado para publicação mas ainda não publicado por um periódico, os autores declaram que receberam autorização do periódico para realizar este depósito.
- O autor submissor declara que todos os autores do manuscrito concordam com a submissão ao SciELO Preprints.