

Documentação Lingüística

Padrões dos Acervos de Línguas Indígenas do Museu

 Sebastian Drude | [My account](#) ▾

Search this site Search

[Informação](#) [Padrões](#) [Procedimentos](#) [Tecnologias](#) [Desafios](#) [Discussão](#) [Ajuda + Wiki](#)

O Portal de Documentação Lingüística (PALIM)

Bem-vinda e bem-vindo ao portal de Documentação Lingüística “PALIM”

— Padrões dos Acervos de Línguas Indígenas do Museu —

Este portal é dedicado à **DOCUMENTAÇÃO LINGÜÍSTICA** — um campo de atuação que se preocupa em criar registros duradouros de línguas e culturas, em particular línguas ameaçadas de desaparecerem dentro de poucas gerações. No Brasil, todas ou quase todas as línguas indígenas se encontram nesta situação.

Para evitar que os registros se percam, e para providenciar o acesso a estes registros para pessoas interessadas e autorizadas, elas podem ser guardadas em **ACERVOS DIGITAIS**. Isto implica no uso de tecnologias avançadas para gravar, preparar e organizar estes registros. Por isto uma grande parte das páginas deste portal se preocupam com estas tecnologias.

Destinatários e Objetivos



O portal se direciona a todas as pessoas interessadas em línguas e culturas indígenas e preocupadas com seu desaparecimento, mas em particular a pessoas que atuam no campo da documentação lingüística (DL): lingüistas, indígenas, indigenistas e simpatizantes. Muitas destas pessoas não têm experiência com tecnologias digitais e de informação, e muitas sofrem em achar e ler as informações existentes sobre isso, pois a maior parte existe somente em Inglês. Por isto a necessidade para este portal, em Português.

Ao mesmo tempo, surgem no Brasil iniciativas para criar acervos digitais de línguas indígenas — especialmente em Museus: No Museu Paraense Emílio Goeldi (do Ministério de Ciência e Tecnologia, MCT, em Belém do Pará), e no Museu do Índio (da FUNAI, no Rio de Janeiro), em cooperação com pesquisadores do Museu Nacional (da Universidade Federal do Rio de Janeiro). Portanto, este portal também quer servir para documentar as soluções que nós achamos, os padrões que adotamos, as metodologias que adotamos, e o progresso que estamos fazendo em criar os acervos.

Se você está procurando informação sobre algum aspecto da Documentação Lingüística, use os menus em cima para navegar e procurar.



Colaboração

Este portal é um WIKI — isto quer dizer que no futuro outras pessoas podem contribuir para completar, corrigir ou comentar as informações, **inclusive você**. Por enquanto, muita informação está disponível em Inglês, em outros *sites*. Estamos procurando pessoas que se dispõem a traduzir estas informações para o Português. Caso que você queira colaborar com isto ou com outros aspectos de completar este portal, por favor entre em contato conosco.

Os links para outros sites não implicam necessariamente que os autores deste portal concordam com eventuais opiniões expressadas nestes outros sites.

page_revision: 15, last_edited: 15 Sep 2009, 21:13 -0-300 (160 days ago)

[Stop watching site palim.wikidot.com](#) [?]

[Edit](#) [Tags](#) [History](#) [Files](#) [Print](#) [Site tools](#) [+ Options](#)

Documentação Lingüística

Padrões dos Acervos de Línguas Indígenas do Museu

Sebastian Drude | My account

Search this site Search

Informação Padrões Procedimentos Tecnologias Desafios Discussão Ajuda + Wiki

Línguas Indígenas

Informação » Línguas Indígenas

Introdução

O Brasil é um país com uma grande diversidade lingüística: falam-se no Brasil não somente o Português Brasileiro e várias outras línguas trazidas por imigrantes, mas também no mínimo **ca. 150 línguas indígenas**. Isto pode parecer muito, e de fato só tem mais ou menos uma dúzia de países no mundo que têm mais línguas. Mas é bem provável que o Brasil, e toda a América do Sul, teve muito mais línguas, talvez até oito vezes mais, 500 anos atrás, na época do primeiro contato com os Europeus.

Muitas vezes se fala em “dialetos” indígenas, o que usualmente implica que estes idiomas seriam menos desenvolvidos, inferiores, menos valiosos do que “línguas verdadeiras”. Isto são preconceitos sem fundamento. Os especialistas em línguas, os lingüistas, sabem muito bem que não existem línguas menos completas ou valiosas que outras — todas as línguas são línguas plenas, com gramáticas e vocabulários elaborados e adaptados ao contexto cultral em que a língua é usada. Na lingüística, o termo “dialeto” designa uma variedade geográfica de uma língua, sem implicações sobre sua complexidade. Assim, o Português falado no Brasil e o Português falado em Portugal são dois dialetos maiores da língua portuguesa, que por sua vez podem ter vários dialetos menores.

Usando estes conceitos, as línguas indígenas não são “dialetos”, mas sim, línguas plenas, independentes, completas. É verdade que algumas delas tenham, por sua vez, dialetos, um pouco diferentes entre si mas ainda próximos o suficiente para que seus falantes se comunicam sem maiores dificuldades e sem ter que aprender o idioma do outro.

A contagem das línguas indígenas no Brasil é complicada e por enquanto não muito confiável, porque ainda não conhecemos bem as línguas no Brasil. Existem várias listas com muitos nomes, mas, em muitos casos, não sabemos por certo se um dado nome é nome de um grupo indígena, um nome de um dialeto (uma variedade geográfica de uma outra língua), ou de uma língua. Também, em geral, é difícil determinar se dois idiomas são variedades da mesma língua ou línguas diferentes. Finalmente, há vários grupos indígenas que vivem sem contato com a sociedade nacional e podem falar outras línguas, ainda desconhecidas.

As línguas são a componente e o veículo mais importante das culturas indígenas — cada língua reflete experiência acumulada da comunidade de seus falantes e é perfeitamente adaptada às suas necessidades ambientais, sociais, culturais, espirituais etc.

Uma página que informa sobre as línguas e sua história está no site do Instituto Socio-Ambiental: [Línguas Indígenas no Site do ISA](#). No mesmo [site do ISA sobre os Povos Indígenas no Brasil](#) há informações sobre todos os povos indígenas do Brasil, inclusive sobre suas línguas.

As seguintes páginas dão uma boa introdução sobre os assuntos mencionados nesta página: [Introdução às línguas indígenas no site do programa ProDoc](#).

A situação das línguas

A grande maioria das línguas indígenas, não só no Brasil, está sob forte pressão. Muitas línguas podem deixar de serem faladas em pouco tempo ou em algumas gerações.

Ver a página [Línguas Ameaçadas](#).

O conhecimento sobre as línguas

A grande maioria das línguas no Brasil (e no mundo) não são bem conhecidas e receberam poucos estudos. Este quadro está começando a mudar com o crescimento da lingüística acadêmica em algumas instituições no Brasil.

Ver o artigo de Denny Moore [1] e, entre outros, este artigo de Lucy Seki [A lingüística indígena no Brasil](#) [4].

A classificação das línguas

As línguas indígenas se agrupam em várias **famílias**, isto é, grupos de línguas irmãs que se desenvolveram durante séculos a partir de uma antiga língua mãe, como se desenvolveram as línguas românicas (Português, Francês, Espanhol, Italiano, Romêno, Catalão, etc.) a partir do Latim. Neste caso fala-se de parentesco genético entre línguas, o que não deve ser confundido com parentesco genético biológico (entre indivíduos e espécies).

Em alguns casos, as famílias lingüísticas por sua vez mostram uma afinidade entre si que sugere que as suas línguas-mãe por sua vez são línguas irmãs, que se desenvolveram de uma língua-mãe mais antga ainda. Por exemplo, as línguas românicas são aparentadas com as línguas germânicas, as línguas eslavas, as línguas celtas e muitas outras famílias. No Brasil, falamos nestes casos de **troncos lingüísticos**, compostos de famílias — como o “tronco indo-europeio” que engloba todas as famílias mencionadas. Em terminologias usadas em outros países, fala-se de uma grande família lingüística com várias sub-famílias ou “galhos”.

Também existem diversas línguas isoladas, isto é, línguas de que não sabemos de nenhuma língua irmã. Conhecendo melhor estas línguas não-classificadas, é capaz que uma ou outra pode ser demonstrada a pertencer a uma ou outra família lingüística.

No Brasil, temos:

- dois troncos lingüísticos (famílias compostas por sub-famílias):
 - Tupi**, com 10 sub-famílias e um total de ca. 37 línguas (no Brasil, no total em todos os países mais que 40), e
 - Macro-Gê**, com 9 sub-famílias e um total de ca. 16 línguas;
- quatro famílias grandes (com 10 línguas ou mais):
 - Arawak**, com ca. 16 línguas (no Brasil, no total em todos os países mais que o 36),
 - Karib**, também com ca. 16 línguas (no Brasil, em todos os países talvez o dobro),
 - Pano**, com ca. 12 línguas, e
 - Tucano**, com ca. 11 línguas;
- seis famílias lingüísticas de tamanho médio (com 3 línguas ou mais):
 - Arawá** (7 línguas),
 - Nadahup ('Maku')** (5 línguas),
 - Chapacura** (5 línguas),
 - Yanomami** (4 línguas),
 - Katukina** (4 línguas),
 - Nambikwara** (3 línguas);
- quatro famílias lingüísticas menores (com só uma ou duas línguas no Brasil):
 - Jabuti** (2 línguas, só no Brasil),
 - Mura-Pirahã** (2 línguas, só no Brasil),
 - Kadiweu** (no Brasil, só uma língua), e
 - Bora** (no Brasil, só uma língua);
- sete línguas isoladas ou nao-classificadas:
 - Aikaná**, **Iranxe**, **Kanoê**, **Kwazá**, **Máku**, **Trumai** e **Ticuna**.

Muitas vezes são publicados listagens com mais línguas. Isto usualmente é causado pelo fato que os autores contaram dialetos maiores como se fossem línguas separadas, apesar de que há indícios que os falantes destes idiomas conseguem se comunicar entre si, cada um usando seu idioma. Nestes casos, nós contamos estas variedades como uma só língua.

Fontes & para saber mais

- [Moore 2008 1] um artigo em Inglês que apresenta e discute a situação das línguas no Brasil, com número de falantes
- [Rodrigues 1986 2] um livro em Português que apresenta as línguas no Brasil e introduz a noções básicas de sua classificação.
- [Fabre 2001ss. 3] um livro com bibliografias bastante completas sobre todas as línguas da América do Sul. Também *online*: <http://butler.cc.tut.fi/~fabre/BookInternetVersio/Alkusivu.html>

Referências

- Moore, Denny (2006). "Brazil: Language Situation". In: Keith Brown (Editor-in-Chief): *Encyclopedia of Language & Linguistics*, Second Edition, volume 2, pp. 117-128. Oxford: Elsevier.
- Rodrigues, Arnon Dall'Igna (1986). *Línguas Brasileiras. Para o Conhecimento das Línguas Indígenas*. Missão Aberta 11. São Paulo: Edições Loyola.
- Fabre, Alain (2001). *Manual de las lenguas indígenas sudamericanas*. LINCOM handbooks in linguistics 4. München & Newcastle: Lincom Europa. *online*: <http://butler.cc.tut.fi/~fabre/BookInternetVersio/Alkusivu.html>
- Seki, Lucy (1999). A lingüística indígena no Brasil. *D.E.L.T.A.* 15: pp. 257-90. *online*: <http://www.scielo.br/pdf/delta/v15nspe/4019.pdf>

page_revision: 12, last_edited: 15 Sep 2009, 16:13 -0-300 (160 days ago)

Stop watching site palim.wikidot.com [?]

Edit Tags History Files Print Site tools + Options

[Help](#) | [Terms of Service](#) | [Privacy](#) | [Report a bug](#) | [Flag as objectionable](#)

Unless otherwise stated, the content of this page is licensed under [Creative Commons Attribution-ShareAlike 3.0 License](#)

Documentação Lingüística

Padrões dos Acervos de Línguas Indígenas do Museu

Sebastian Drude | My account

Search this site Search

Informação Padrões Procedimentos Tecnologias Desafios Discussão Ajuda + Wiki

Línguas Ameaçadas

Informação » Línguas Ameaçadas

Introdução

Línguas desaparecem quando ninguém mais as usa. Muitas línguas que já foram faladas no passado não são faladas mais.

Quando pensamos nisto, talvez lembramos do Latim, uma língua falada antigamente (há uns quinze a vinte e cinco séculos) pelos romanos. Hoje ainda há pessoas que sabem usar o Latim, especialmente pessoas da Igreja Católica. O Latim até hoje até é a língua oficial do Vaticano (o menor estado do mundo). Mas o Latim não morreu sem deixar uma grande herança: as línguas românicas, que se desenvolveram a partir do Latim Vulgar, o Latim falado pela população simples no grande reino romano.

De fato, é normal que as línguas se transformem, não existe língua que seja igual hoje ao que era 500 anos atrás.

Mas também há muitos casos de línguas que deixam de ser faladas sem se transformar em outras. isto pode acontecer porque todos os falantes morrem, por exemplo em uma catástrofe natural, ou numa guerra ou até num genocídio.

Mas é muito mais comum que línguas desapareçam porque são **abandonadas** — os falantes da língua deixam de falar ela em todos os contextos ("domínios"), e mudam para uma outra língua, a **língua dominante** ou **língua-alvo**. A situação piora substancialmente quando em algum momento os falantes adultos deixam de transmitir a língua ameaçada para a próxima geração: eles não mais ensinam a língua para seus filhos. Uma língua que está em uma situação de pressão, em que isto pode facilmente acontecer num futuro próximo, ela é uma **língua ameaçada**.

Quando já não tem mais crianças que aprendem esta língua como língua materna, a língua é **moribunda** ("morrendo", ou "altamente ameaçada" — pois se nada acontece, ela morre junto com os últimos falantes das gerações mais velhas). Se ela já não é mais usada no cotidiano, ela de fato já morreu (é uma **língua morta**, mesmo que ainda tenha pessoas que se lembram dela ou poderiam usá-la). Quando não tem mais nem pessoas que se lembram dela (mais do que algumas poucas palavras), ela é uma **língua extinta**.¹

Com a expansão da cultura europeia em várias partes do mundo e alguns outros impérios regionais (como na China), na sua forma tradicional chamado **colonialismo** e na forma moderna chamada **globalização** (no nível nacional: **integração nacional**), o fenômeno da extinção de línguas ficou cada vez mais freqüente. Isto não vale somente para o Brasil, mas para todas as regiões do mundo: em quase todos os países a maioria das línguas é ameaçada, e em muitos uma grande parte das línguas já é moribunda.

Em um cálculo estatístico, Rodrigues estima que no Brasil existiam até 1275 línguas indígenas por volta de 1500, quando da chegada dos portugueses na costa brasileira.[1] Mesmo se esta cifra pode parecer exagerada, é óbvio que o Brasil perdeu a maior parte de sua grande diversidade lingüística nos séculos passados. Muitas destas línguas morreram sem que nós soubéssemos mais do que um nome, ou nem isso. E o que é mais grave: se nada acontece, é bem capaz que daqui a 100 ou 150 anos todas ou quase todas as línguas indígenas desapareça também.

Como o gráfico acima mostra, o tamanho mediano das línguas indígenas faladas no Brasil é por volta de 200 falantes, isto é, a metade das línguas tem mais e metade tem menos que 200 falantes. Mesmo que as cifras absolutas das populações dos grupos indígenas esteja aumentando, muitas vezes as línguas estão ameaçadas de desaparecer.

Defendemos que não devemos ficar indiferentes diante deste desenvolvimento de dimensões catastróficas. Especialmente as comunidades dos falantes e seus descendentes sentem os danos irreparáveis. Nas palavras de Bruna Franchetto:

A agonia e a morte de uma língua resulta em sérios danos à saúde intelectual de seu povo, a suas tradições orais, suas formas artísticas (poesia, música, oratória), seu conhecimento, sua ontologia e perspectiva cosmológica.

A comunidade mundial dos lingüistas se deu conta sistematicamente de que as línguas estão desaparecendo cada vez mais rapidamente somente no início dos anos noventa do século XX. A **documentação lingüística** surgiu neste contexto: pelo menos podemos tentar documentar algo destas línguas para preservar sua memória para gerações futuras, e talvez contribuimos em alguns casos com isto para sua manutenção.

Fatores

Pode ter várias razões para isto, mas usualmente há alguns fatores que quase sempre estão presentes quando isto acontece:

- minoritária**: a língua ameaçada é falada por uma minoria em um país ou uma região
- integração nacional / globalização**: o contato da comunidade da língua ameaçada com a sociedade envolvente está se intensificando
- transformação cultural**: muitas vezes, a cultura tradicional está rapidamente integrando elementos da sociedade envolvente, e abandonando elementos da cultura tradicional
- diglossia**: a maioria da população daquele país ou daquela região usa uma outra língua, que é **dominante** em comparação com a língua ameaçada:
 - as pessoas que usam a língua dominante têm mais poder econômico
 - a língua dominante é muito mais usada em domínios importantes, como educação, comunicação escrita, e mídias de massa, comércio etc.
 - a língua dominante tem mais prestígio
- percepção de inferioridade**: os falantes da língua ameaçada sofrem de baixa auto-estima e relacionam isto à sua língua, ou creem que sua língua não é tão apta para o uso nos domínios ocupados pela língua dominante
- migração**: em vários casos, os falantes da língua ameaçada se mudam para outros lugares, por exemplos cidades onde não há muitos falantes da língua ameaçada
- abandono**: os falantes querem que seus filhos tenham melhores chances, e acreditam que a língua ameaçada poderia atrapalhá-los

Determinar o grau de ameaça

Vários indicadores foram propostos para determinar se e em que grau uma língua é ameaçada. Ver um documento da UNESCO sobre este assunto [2].

- Número absoluto de falantes**: Línguas com poucos falantes têm mais dificuldades para se manter e usualmente sofrem uma pressão mais forte de línguas dominantes. Porém, tem casos de línguas com poucas dúzias que se mantiveram por décadas (graças a um contexto cultural intacto e um alto grau de isolamento da sociedade envolvente), e há outros casos, por exemplo na Índia e na China ou nas Filipinas, em que línguas com dezenas de milhares ou mais de falantes estão sendo abandonadas por uma geração inteira.
- Numero relativo de falantes**: que menor a comunidade da língua ameaçada em relação à sociedade envolvente, que mais difícil é mantê-la. A questão é qual é o parâmetro para a comparação
- Transmissão intergeracional**: provavelmente o fator decisivo: uma língua que é aprendida por poucas crianças (ou nenhuma) é altamente ameaçada
- Mudanças culturais e sociais**: transformações rápidas e drásticas podem diminuir ou tirar os contextos em que a língua ameaçada é necessária ou funcional
- Respostas** a novas esferas de uso, mídia: línguas que se adaptam a novos contextos de uso, elaborando seu vocabulário e se adaptando ao uso escrito, inclusive em novas mídias, têm mais chances de se manterem do que se em novos contextos se usa somente a língua dominante
- Materiais pedagógicos**: se há alfabetização e escolarização básica na língua ameaçada, ela tem mais chances que ela não é completamente abandonada pelos jovens falantes

Bilingüismo

A longo prazo, as chances de sobreviverem para muitas línguas está num **bilingüismo estável**.

Diferentemente de que muitos brasileiros pensam, bilingüismo (o fato de uma pessoa ou um grupo de pessoas dominar duas línguas com quase igual fluência, como maternas) ou multilingüismo (se mais línguas são envolvidas) não prejudicam o desenvolvimento de uma pessoa. No mundo atual e na história, multilingüismo de fato é mais comum do que o monolingüismo, que caracteriza as sociedades dos grandes estados, principalmente na Europa e em grandes partes das Américas.

Se as duas línguas são aprendidas plenamente e a pessoa consegue lidar naturalmente com seu uso em ambientes e situações diferentes, o bilingüismo provavelmente até ajuda o desenvolvimento intelectual da pessoa e fortalece sua competência social. O bilingüismo é **estável** se duas línguas convivem em uma sociedade por várias gerações, possivelmente apesar de uma relação de diglossia (em que uma língua é dominante em relação à outra, *ver acima*). Em vários países há exemplos de um bilingüismo estável, mas nas últimas décadas mesmo línguas que costumaram exemplos positivos, como o Friso ou o Sorbo na Alemanha, sofram um maior grau de ameaça com a globalização. É irônico que a mesma modernização tecnológica, que é um fator crucial na globalização, ao mesmo tempo cria as condições para uma documentação moderna, em forma de acervos digitais duradouros.

O que fazer? : Respostas

O que precisa ser feito a nível nacional:

- Conhecer melhor a situação das línguas – levantamento (critérios UNESCO), sua identificação (há planos de incluir a língua no próximo censo brasileiro)
- Reconhecer as línguas como patrimônio e assumir a tarefa de protegê-las - há projetos pilotos para um levantamento nacional das línguas indígenas organizado pelo IPHAN – tem um Seminário Legislativo Sobre a Criação do Livro de Registro das Línguas
- Apoiar a pesquisa e centros regionais
- Investir na formação de lingüistas, inclusive de lingüistas indígenas (cf. Maria Pankararu, Rogério Ferreira da Silva)

- Conscientizar os falantes das línguas indígenas e outros sobre a situação e discutir opções
- Auxiliar o desenvolvimento de uma política lingüística:
 - Ajudar estabelecer a língua na educação
 - Desenvolver a escrita, materiais
 - Identificar domínios a serem (re)tomadas
- Documentar as línguas e culturas > daí o investimento na [documentação lingüística](#) e a criação [deste portal](#)...

Bibliography

1. Rodrigues, Arnon Dall'Igna (1993). "Línguas Indígenas: 500 anos de descobertas e perdas". *Ciência Hoje* 16/No 95: pp.20-26.
2. UNESCO Ad hoc Expert Group on Endangered Languages, (Matthias Brenzinger et al.) (2002). *Language Vitality and Endangerment*. online: <http://www.unesco.org/culture/ich/doc/src/00120-EN.pdf>
3. *Wikipedia*: http://en.wikipedia.org/wiki/Endangered_language

Notas

1. Esta terminologia, que usa metáforas como se línguas fossem organismos ou espécies biológicas, não é ideal e foi criticada por vários especialistas. Termos melhores podem ser desenvolvidos e introduzidos. Também há especialistas que usam definições ligeiramente diferentes — por exemplo, para vários uma língua é "ameaçada" quando não tem mais crianças aprendendo ela (o que aqui chamamos de "moribunda"). Na concepção defendida aqui, isto cria o risco que as pessoas subestimam o perigo para as línguas que estão sob uma forte pressão de serem abandonadas ("ameaçadas", na terminologia aqui usada), mas que ainda (por enquanto) têm crianças aprendendo ela. Também poucos especialistas fazem a diferença entre uma "língua morta" e uma "língua extinta" que propomos aqui.

page_revision: 10, last_edited: 15 Sep 2009, 19:59 -0-300 (160 days ago)
 Stop watching site palim.wikidot.com [?]

[Edit](#) [Tags](#) [History](#) [Files](#) [Print](#) [Site tools](#) [+ Options](#)

Documentação Lingüística

Padrões dos Acervos de Línguas Indígenas do Museu

Informação
Ajuda + Wiki

Padrões

Procedimentos

Tecnologias

Desafios

Discussão

 Sebastian Drude | [My account](#)

Search this site Search

Documentação Lingüística

[Informação](#) » Documentação Lingüística

Introdução

Documentação lingüística é um campo de atuação da lingüística que se ocupa com a criação de registros duradouros de línguas em seu uso, usualmente através da criação de acervos digitais que permitem o acesso e uso deste material mesmo por pessoas que não falam e entendem a língua em questão.

A documentação lingüística neste sentido surgiu por volta do ano 2000 no contexto da consciência cada vez maior de que a grande maioria das línguas do mundo são [línguas ameaçadas](#).

Documentação, neste sentido, não deve ser confundida com, e é complementar e interligado à *descrição*, que se ocupa com a análise da estrutura de uma língua.

Há um trabalho de Nikolaus Himmelmann que é muitas vezes citado como o passo principal na fundação deste novo ramo.^[1]

Panorama geral

- Uma documentação moderna contém um corpus de gravações em áudio e vídeo (**dados primários**) de eventos culturais e em particular de eventos de fala natural, isto é, de "textos".
- Para documentar o uso da língua em toda amplitude e diversidade, e para garantir a utilidade para diversos objetivos posteriores que possam querer fazer uso da documentação, é preferível que os textos pertençam a uma grande variedade de gêneros e domínios.
- Os textos usualmente estão acompanhados por anotações (**dados secundários**). O mínimo de anotações em DOBES são uma transcrição e uma tradução, mas pode ter muitas outras anotações (glosas, indicações categoriais das unidades, etc.).
- Este material, os dados primários e secundários, devem ser catalogados com **meta-dados**, e organizados conforme seu conteúdo.
- Uma documentação hoje em dia deve ser digital em formatos estandarizados e permanentes, e acessível ao grande público, p.ex. via a internet.

Projetos e programas de documentação lingüística

Existem diversas iniciativas para desenvolver e aplicar a documentação lingüística. Ver a página [Projetos Afins](#) para descrições sumárias e *links*.

Publicações

Até agora, quase todas as publicações sobre documentação lingüística são em inglês. Com a ajuda dos especialistas Brasileiros, esperamos providenciar neste portal informações básicas em português.

Entre as publicações mais relevantes em Inglês e Espanhol há:

- Um **manual** que dá um excelente panorama geral da documentação lingüística, escrito e editado por membros do do programa DOBES: ***Essentials of Language Documentation*** organizado por Jost Gippert, Nikolaus P. Himmelmann e Ulrike Mosel and publicado em sua primeira edição (em Inglês) por Mouton - de Gruyter, 2006. A segunda edição é em Espanhol: ***Bases de la documentación lingüística***, John B. Haviland & José Antonio Flores Farfán (coordinadores da versão em espanhol), México: Instituto Nacional de Lenguas Indígenas, 2007. Algumas informações e materiais adicionais a este livro se encontram nas [páginas do livro //Essentials of Language Documentation//](#).
- Uma série de "Working Papers" editados pelo HRELDLP / SOAS: Peter K. Austin, (editor) [Language Documentation and Description](#).
- Uma revista científica: [Language Documentation & Conservation \(LD&C\)](#), editado como revista digital na University of Hawai'i Press, desde 2007.

Este portal

Este portal quer ajudar estabelecer e fortalecer a Documentação Lingüística no Brasil, difundindo conhecimentos entre pessoas interessadas em participar no objetivo como identificado acima: lingüistas, indigenistas, e em particular membros e ativistas de grupos indígenas.

O foco deste portal é em aspectos tecnológicos da Documentação Lingüística, porque uma grande parte das atividades fazem uso de tecnologias modernas. Isto é um obstáculo para muitas pessoas potencialmente interessadas em participar na documentação de uma língua.

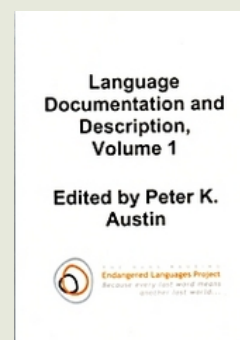
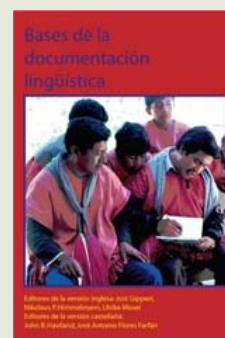
Para ver os assuntos principais a serem abordados, ver a página [sobre este Wiki](#).

Referências

1. Himmelmann, Nikolaus P. (1998). "Documentary and descriptive linguistics". *Linguistics* 36: pp.161-95. Uma versão mais completa é disponível *online*: <http://www.hrelp.org/events/workshops/elldp2005/reading/himmelmann.pdf>

page_revision: 8, last_edited: 16 Sep 2009, 12:18 -0-300 (159 days ago)
Stop watching site palim.wikidot.com [?]

Edit Tags History Files Print Site tools + Options



Documentação Lingüística

Padrões dos Acervos de Línguas Indígenas do Museu

Informação
Ajuda + Wiki

Padrões

Procedimentos

Tecnologias

Desafios

Discussão

 Sebastian Drude | [My account](#) ▾

Search this site

Search

Projetos afins

[Informação](#) » [Documentação Lingüística](#) » [Projetos afins](#)

Introdução

Desde o início dos anos 90, e em particular desde 2000, surgem várias iniciativas que se preocupam com [línguas ameaçadas](#) e a [documentação lingüística](#). A maioria destas iniciativas mantem *web-sites*, usualmente em Inglês.

Iniciativa Brasileira

Em 2009 foi lançado o [ProDocLin](#) ("Programa de Documentação de Línguas e Culturas Indígenas"), pelo [Museu do Índio](#) / FUNAI / Ministério de Justiça / Governo Federal do Brasil. O programa ProDocLin é parte de uma iniciativa maior, ProDoc, que também quer documentar culturas indígenas. Atualmente, tem os seguintes projetos neste programa:

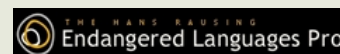


- Apiaká;
- Arara/ikpeng;
- Desano;
- Kanoê;
- Kawaiwete/Kaiabi;
- Kisêdjê;
- Paresi/Haliti;
- Rikbaktsa;
- Shawâdawa;
- Yawanawa

Iniciativas internacionais

Há algumas iniciativas com atuação e significância mundial.

- [DOBES](#) ("Documentação de Línguas Ameaçadas", em Alemão: "DOkumentation BEdrohter Sprachen"), lançada em 2000 pela Fundação Alemã "Volkswagenstiftung" (que não pertence à empresa do mesmo nome). Foi crucial no estabelecimento da nova área, e até hoje apoia projetos de documentação de línguas individuais. No programa DOBES houve 5 projetos no Brasil, os primeiros 3, no Alto Xingú, já participaram da fase piloto:
 - Kuikuro (Karib, Bruna Franchetto, Museu Nacional / UFRJ)
 - Trumai (isolada, Raquel Guirardello, MPI Nijmegen & Museu Goeldi)
 - Awetí (Tupí, Sebastian Drude, Freie Univ. Berlin & Museu Goeldi)
 - Mawé (Tupí), Bakairí, Kashuyana (Karib) (Sérgio Meira, Leiden & Museu Goeldi)
 - Cashinahua (Pano, E. Camargo etc., Paris X)
- [HRELP](#) ("The (Hans Rausing) Endangered Languages Project"), da "Escola de Estudos Orintais e Africanos", Inglês: "School of Oriental and African Studies". Com apoio do ELDP houve/há 7 projetos no Brasil ou com participação de Brasileiros:
 - Apurinã (Aruak, Sidney Facundes, UFPA)
 - 4 línguas Tupí (D. Moore etc, Museu P.E. Goeldi)
 - Línguas do Chaco (V. Grondona & F. Sândalo, UNICAMP)
 - Ofayê (Macro-Jê, Eduardo Ribeiro, UFGO)
 - Karo (Tupí, Nilson Gabas Jr., M.P.E. Goeldi)
 - Enawene-Nawe (Aruak, Ubiray Rezende, UFRJ)
 - Waikhana + Wanano (Tukano Oriental, K. Stenzel, UFRJ)
- [EMELD](#) ("Electronic Metastructure for Endangered Languages Data"), da "LINGUIST-List" (Estados Unidos). Este projeto, ativo nos anos 2001 a 2007, não financiou projetos de documentação mas contribuiu significativamente para o desenvolvimento de padrões e procedimentos, metodologias e tecnologias na área da documentação lingüística.
- [DEL](#), ("Documenting Endangered Languages") da National Science Foundation (Estados Unidos). Ativo desde 2006.



Iniciativas regionais e de outros países

- AILLA (Univ. Texas, USA)
- PARADISEC (U. Sydney etc., Austrália)
- LACITO (CNRS, França)
- LDA (LDC, U. Pennsylvania)
- ELF (Yale U.)
- Rosetta Project (Stanford U.)
- Alaska Native Language Center (Fairbanks)
- GFBS, Alemanha

page_revision: 15, last_edited: 16 Sep 2009, 11:44 -0-300 (159 days ago)
Stop watching site palim.wikidot.com [?]

[Edit](#) [Tags](#) [History](#) [Files](#) [Print](#) [Site tools](#) [+ Options](#)

Esquema Geral

Informação » Esquema Geral

Todos os Padrões e Procedimentos são aqui organizados e visualizados em um esquema geral. Por favor clicar na imagem para ver o esquema no tamanho normal.

As explicações em baixo podem ser escondidas ou mostradas clicando no texto "clique aqui...".

Explicações

Geral

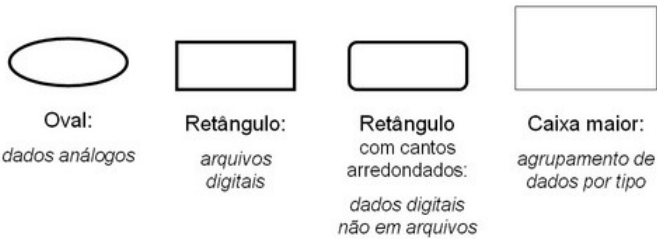
- clique aqui para esconder explicações gerais

- Cada objeto de dados (por exemplo, cada gravação, cada foto, cada nota ou anotação) é representado por um ícone com um rótulo
- Para se informar sobre cada tipo de objeto, veja a lista de abreviações em **Padrões**
- Os dados "brutos" (não processados) ficam na coluna à esquerda
- Os dados processados estão mais à direita
- Os dados definitivos, isto é, arquivos digitais a serem arquivados, têm cores sólidas
- As setas simbolizam procedimentos de processamento, transformação, relações entre os dados e arquivos, também com um rótulo único
- Para se informar sobre cada tipo de procedimento, veja a lista de abreviações em **Procedimentos**

- clique aqui para esconder explicações gerais

Ícones : objetos, formatos de dados e de arquivos

- clique aqui para esconder explicações dos ícones



- Caixas ovais: dados análogos, anotações em papel
- Caixas retangulares: arquivos digitais
- Caixas retang. arredondadas: dados digitais, outros recipientes que arquivos
- Caixas maiores: agrupações por tipos de dados (brutos)

- clique aqui para esconder explicações dos ícones

Cores: tipos de dados

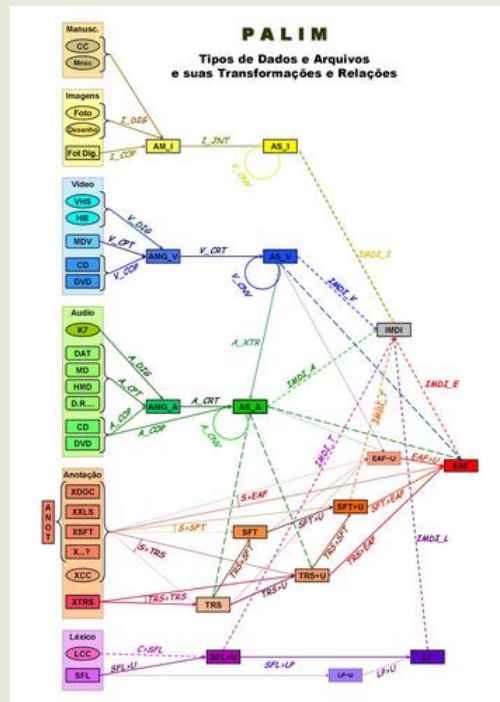
- clique aqui para esconder explicações das cores



- Amarelo e bege: imagem, desenho, manuscrito em papel
- Azul: vídeo
- Verde : áudio
- Laranja / vermelho / rosado : anotação
- Lilás : léxico
- Cinza : meta-dados

- clique aqui para esconder explicações das cores

Setas : procedimentos, transformações, relações (links)





- Setas sólidas: conversão integral da informação
- Setas pontuadas: serve-como-fundamento-para, datilografia
- Setas interrompidas: *'links'*, referências
- Setas finas: procedimentos / relações não recomendadas

- clique aqui para esconder explicações das setas

Elementos recorrentes dos rótulos de objetos

Para ver o significado de cada rótulo, veja a lista de abreviações em Padrões.

- clique aqui para esconder explicações dos rótulos

| Elemento | Significado |
|----------|--|
| A | Arquivo |
| AM | Arquivo de Mídia |
| AMG | Arquivo de Mídia Global |
| AS | Arquivo de Sessão (a ser arquivado) |
| TRS | Transcrição em formato de <i>Transcriber</i> |
| SF | Standard Format (texto com marcadores de campo) |
| SFT | Standard Format Texto (anotação de textos, transcrição, tradução...) |
| SFL | Standard Format Léxico |
| _I | Imagem |
| _V | Vídeo |
| _A | Aúdio |
| +U | Unicode |
| -U | sem ser Unicode |

- clique aqui para esconder explicações dos rótulos

Elementos recorrentes das rotulos de setas

Para ver o significado de cada seta, veja a lista de abreviações em Procedimentos.

- clique aqui para esconder explicações das setas

| Elemento | Significado |
|--|--|
| ..._... (<i>traço inferior</i>) | Procedimento em geral |
| C_ | Criação (gravação) de dados |
| IMDI_ | Inclusão nos metadados IMDI |
| ...X...>...Y... (<i>maior</i>) | Conversão de dados X em dados Y |
| S> | Segmentar (<i>time linking</i>) |
| A | Relacionado a dados de áudio |
| I | Relacionado a dados de imagem/manuscrito |
| L | Relacionado a dados de vídeo |
| T | Relacionado a um projeto toolbox |
| V | Relacionado a dados de vídeo |
| CNV | Converter |
| COP | Copiar |
| CRT | Cortar |
| DIG | Digitalizar |
| Outras abreviações são as mesmas que usadas nos <u>rótulos</u> . | |

- clique aqui para esconder explicações das setas

Documentação Lingüística

Padrões dos Acervos de Línguas Indígenas do Museu

 Sebastian Drude | [My account](#) 

Search this site

[Informação](#) [Padrões](#) [Procedimentos](#) [Tecnologias](#) [Desafios](#) [Discussão](#) [Ajuda + Wiki](#)

Padrões : Observações Gerais

[Padrões](#) » [Padrões : Observações Gerais](#)

fold

Table of Contents

- [Introdução](#)
- [Porque um acervo digital](#)
- [Porque padrões e formatos](#)
- [Formatos gerais e específicos, por tipo de dados](#)
 - [Dados primários](#)
 - [Dados secundários](#)
 - [Metadados](#)
- [Dados analógicos versus dados digitais](#)
- [Formatos e encodificações de arquivos de textos](#)

Para ver todas as páginas desta categoria, vai para a página [Padrões](#).

O acesso direto também pode ser feito pela [lista dos rótulos](#) (abreviações) usados no [esquema geral](#).

Introdução

Estas páginas tratam de formatos e padrões de **"dados"** — isto quer dizer, qualquer objeto ou entidade que se quer incluir em alguma forma num acervo digital para sua preservação.

Todos os dados pertencem a certos **tipos** (em particular: imagem, vídeo, áudio, anotação, léxico, metadados) e ocorrem em certos **formatos (genéricos)** (gravação analógica de áudio, arquivo digital de vídeo, arquivo de texto, etc.).

No caso de arquivos digitais, os **formatos específicos** (por exemplo: arquivo de documento no formato *Word*, *PDF* etc.) ainda podem variar por fatores como a **codificação** (por exemplo arquivos de vídeo no formato *MPG* / *MPEG* / *MP4*, com codificação *PAL*, *NTSC*, etc., textos em formato *standard format*, *XML*, etc., com codificação *Unicode* ou *ANSII*, etc....

Os dados ainda podem estar em diferentes **suportes** ou **mídia física**.

Porque um acervo digital

Para não correr o risco de perder valiosos dados documentando eventos importantes, é essencial que eles sejam arquivados em um lugar central.

Para que dados possam ser arquivados, eles precisam ser transformados em arquivos digitais aptas para serem inseridos no acervo digital (servidor [LAT](#)).

A questão aqui é a **sustentabilidade** dos dados, isto é, as chances que os dados ainda são acessíveis em décadas futuras.

Há riscos de vários tipos aqui. Os **suportes físicos** podem perder sua funcionalidade — fitas K7 podem se embaralhar, elas e disquetes podem se desmagnetizar, CDs e DVDs podem ser arranhados ou sofrer deterioração dos materiais, discos rígidos podem quebrar, etc. Um acervo central é um passo importante rumo a uma solução para estes problemas — os dados de um acervo podem facilmente ser copiados (criando cópias de segurança), discos rígidos podem ser trocados, etc. Um acervo central num servidor ao mesmo tempo pode garantir a acessibilidade dos dados a partir de lugares remotos, e cópias parciais para divulgação em lugares sem acesso à rede.

Porque padrões e formatos

Mesmo se a integridade dos dados nos suportes físicos seja garantida, ainda há a questão dos **formatos** — muitos dados só podem ser acessados através de uma certa tecnologia, em particular usando um determinado programa. Estes programas e os formatos criados e lidos por eles mudam com o tempo, e os dados se tornam inacessíveis ou inutilizáveis, mesmo que a sequência digital pode ser lida. A solução para este problema são **padrões (standards) abertos**, isto é, formatos que podem ser entendidos sem precisar der acesso a certos patentes ou programas. Com as especificações publicamente expostos, sempre será possível criar tecnologias que decifram os dados e os transformam em formatos então atuais.

Por isso, é importante que os dados a serem "depositados" (guardados) no acervo obedecam a certos critérios e pertençam a formatos padronizados.

É o objetivo destas páginas identificar e descrever os formatos possíveis e existentes, em particular definir os formatos aceitos para arquivação.

Os padrões de formatos podem ser classificados pelos diferentes critérios mencionados acima.

Formatos gerais e específicos, por tipo de dados

Nós distinguimos entre dados primários (registros diretos de eventos de fala ou de elementos culturais da comunidade de falantes, ou objetos físicos produzidos pelos falantes), dados secundários (representações escritas de dados primários) e meta-dados (representações de propriedades de dados primários e secundários não diretamente ligadas à estrutura da língua).

Por poder ser processado de forma semelhante, abordamos aqui também os cadernos de campo como se fossem imagens (dados primários), apesar de que eles contém usualmente análise e representações escritas de eventos de fala (ou seja, estritamente falando se trata de dados secundários).

Igualmente subsumimos bancos de dados lexicais e descrições aos dados secundários (estricamente, estes tipos de dados são mais abstratos e usualmente não diretamente se referem a um evento de fala ou outros dados primários).

Para todos os dados é importante observar a divisão entre dados analógicos e dados digitais.

No caso de documentos de texto, é bom saber sobre o formato genérico XML, e sobre a codificação UNICODE.

Dados primários

- Imagem / manuscrito
- Vídeo
- Áudio — Sobre padrões e formatos de dados de áudio

Dados secundários

- Anotação
- Léxico
- Descrições e análises

Metadados

- Metadados sobre suportes
- Metadados sobre arquivos
- Metadados sobre sessões
- Metadados gerenciais e estruturais

Dados analógicos versus dados digitais

Uma divisão fundamental distingue dados analógicos e dados digitais.

Dados analógicos representam propriedades de objetos e eventos reais em alguma estrutura física escalável — por exemplo uma intensidade de tinta (ou pigmentos de cor em uma fotografia) em papel, magnetismo em uma fita, etc.

Dados digitais são uma representação mais abstrata — no mundo digital só existem "bits", sendo que cada bit pode ser um "sim" ou um "não", ou então "1" ou um "0"; tudo é composto por sequências bits. Assim, letras são representados por sequências de poucos bits — sete bits são o suficiente para representar um alfabeto de 128 letras e símbolos, cada letra ou símbolo representado por uma sequência diferente de seven bits. Para um alfabeto maior se precisam mais bits.

Igualmente dados de imagens, de audio e de vídeo podem ser "digitalizados" — convertidos de uma representação análogo para uma sequência de bits.

Formatos e encodificações de arquivos de textos

- XML
- UNICODE

Documentação Lingüística

Padrões dos Acervos de Línguas Indígenas do Museu

Sebastian Drude | My account

Search this site Search

[Informação](#)
[Padrões](#)
[Procedimentos](#)
[Tecnologias](#)
[Desafios](#)
[Discussão](#)
[Ajuda + Wiki](#)

Padrões de Metadados

Fold

Table of Contents

[Sobre Metadados](#)
[O padrão IMDI](#)
COLUNAS:
[As categorias das fichas IMDI](#)

Sumário rápido: Padrão recomendado para medadados

O formato recomendado para arquivos de metadados que vão integrar um acervo digital é:

- arquivos XML usando os esquemas IMDI (obrigatório em servidores LAT);
- ou arquivos XML usando o esquema OLAC.

Sobre Medadados

Metadados, ou Metadata, em geral são dados sobre dados. No caso de documentação lingüística, os metadados servem para catalogizar todos os outros tipos de dados, como dados primários (gravações de áudio e vídeo, imagens, etc.) e secundários (anotação). Os metadados especificam as circunstâncias das gravações, quem gravou e quem foi gravado, as línguas usadas, o conteúdo ou assunto, o projeto que financiou o apoiou, etc.

Nos acervos de LAT (Language Archiving Technology), como os no Museu do Índio, no Rio de Janeiro, e no Museu Goeldi, em Belém, há um padrão estabelecido para metadados: o padrão da IMDI (ISLE MetaData Initiative, sendo que ISLE significa por sua vez 'International Standard in Language Engineering').

Este formato permite a organização de vários arquivos (multi-mídia e de anotação) em SESSÕES, e a agrupação de sessões em hierarquias que podem ser procuradas ou navegadas.

O padrão IMDI

O padrão IMDI foi desenvolvido no Instituto Max Planck para Psicolingüística em Nimega em cooperação com outros centros, por volta do dos anos 2001-2003. Arquivos de Meta-dados IMDI são arquivos XML em um formato específico. O formato é definido e descrito neste [documento XSD](#).

Recomendamos este padrão porque a tecnologia LAT, instalada no servidor do Museu Goeldi e no Museu do Índio, se baseia neste formato. Também, este formulário é o mais completo que atualmente existe para o tipo de material que interessa para a documentação lingüística.

Um arquivo de metadados IMDI é como uma grande ficha com muitos campos para diferentes tipos de informação. Ele pode ser preenchido com o uso do programa IMDI-Editor, disponibilizado de graça pelo MPI.

Nas páginas dos links que seguem, indicamos para cada seção (máscara / folha) quais campos do formato devem ou podem ser preenchidos, com exemplos. Cada elemento é descrito em forma de uma linha de uma tabela.

Estas tabelas explicam como fazemos uso do esquema IMDI para sessões

COLUNAS:

| Coluna | Explicação | Carater | Valores |
|-----------------|-------------------------------|---------------|----------------------------------|
| Campo | Nome do campo no editor IMDI | livre | |
| Uso | obrigatório, recomendado etc. | lista fechada | + mostre valores |
| Providen | quem providencia este campo | lista fechada | + mostre valores |
| Repete | O campo pode repetir? | lista fechada | + mostre valores |
| Coment | Comentário, explicação | livre | |
| Exemplo | Valores exemplares | livre | |

As categorias das fichas IMDI

1. [Sessão \(Session\)](#)
2. [Projeto \(Project\)](#)
3. [Conteúdo \(Content\)](#)
4. [Conteúdo: Ator \(Actor\)](#)
5. [Conteúdo & Ator: Língua \(Language\)](#)
6. [Recursos \(Resources\)](#)
7. [Recursos: Arq. de Mídia \(Mediafile\)](#)
8. [Recursos: Arq. de Anotação \(Written Resource\)](#)
9. [Recursos: Fonte \(Source\)](#)

page_revision: 23, last_edited: 12 Jan 2010, 10:13 -0-300 (41 days ago)
 Stop watching site palim.wikidot.com [?]

[Edit](#) [Tags](#) [History](#) [Files](#) [Print](#) [Site tools](#) [+ Options](#)