

Advances in Documentary Linguistics in Brazil

Towards identifying and documenting
Brazil's linguistic diversity

Sebastian Drude, Denny Moore, Rosileide G. Costa
DOBES-Meeting October 2009, Nijmegen

1

Topics

1. Intl. Documentation Projects in Brazil
2. Brazilian National Doc. Prog.: ProDocLin
3. LAT archives in Rio and Belém
4. The PALIM - Portal (Wiki)
5. Brazilian national language survey
6. Languages in 2010 Brazilian census
7. Identification and classification of Brazilian Languages

June 2008

DOBES - file management

2

1 Intl. Documentation Projects

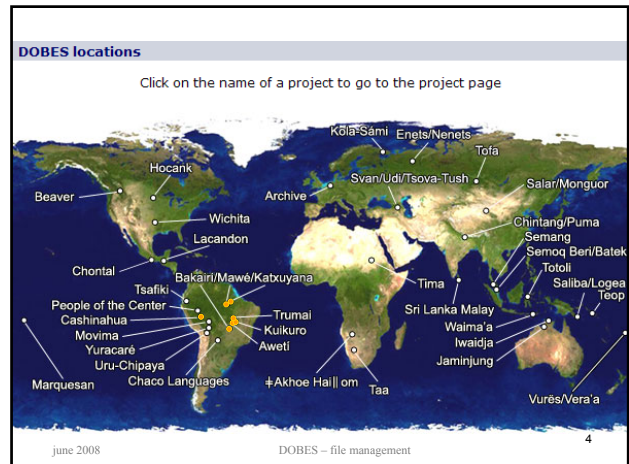
5 DOBES projects — the first 3, in the *Upper Xingu*, participated in the DOBES pilot phase ~2001:

- **Kuikuro** (Karib, by Bruna Franchetto et.al., Museu Nacional / UFRJ)
- **Trumai** (isolated Ig., by Raquel Guirardello et.al., MPI Nijmegen & Museu Goeldi)
- **Awetí** (Tupí, by Sebastian Drude et.al., Freie Univ. Berlin & Museu Goeldi)
- **Mawé** (Tupí), **Bakairí**, **Kashuyana** (Karib) (by Sérgio Meira, U. Leiden & Museu Goeldi)
- **Cashinahua** (Pano, by Eliane Camargo et.al., U. Paris X & UNICAMP)

June 2008

DOBES - file management

3



4

1 Intl. Documentation Projects

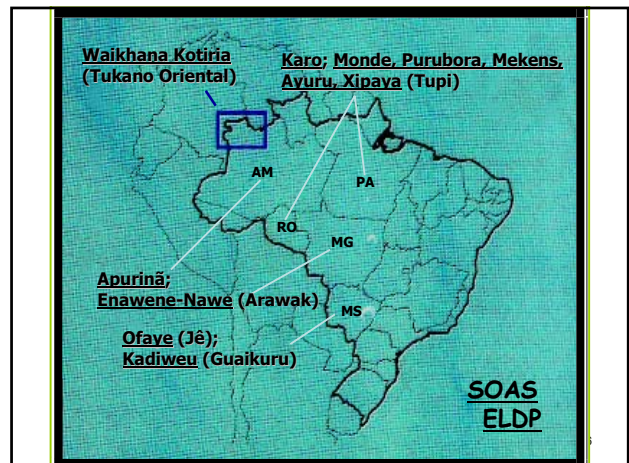
7 **ELDP/SOAS** projects :

- **Apurinã** (Arawak, Sidney Facundes, UFPA)
- **4 Tupí Igs.** (Denny Moore & al, Museu Goeldi)
- **2 Igs. of the Chaco** (Guaicurú + Jê; Ver. Grondona & Fil. Sândalo, UNICAMP)
- **Ofayê** (Macro-Jê, Eduardo Ribeiro, UFGO)
- **Karo** (Tupí, Nilson Gabas Jr., Museu Goeldi)
- **Enawene-Nawê** (Arawak, Ubiray Rezende, UFRJ)
- **Waikhana + Wanano** (Tukano Oriental, Kristine Stenzel, UFRJ)

June 2008

DOBES - file management

5



2 Brazilian National Prog.: ProDocLin

- In August 2008, a 3-year pilot program was approved by the Brazilian government via the FUNAI (Indigenous Agency, Min. o. Justice)
- Provoked envy and intense criticism from some opponents, but went ahead
- ProDocLin adapts the DOBES methodology and uses LAT technology
- Data is to be included in the archives at Museu do Índio, possibly Museu Goeldi
- Coordination: Bruna Franchetto (Museu Nacional, DOBES-Kuikuro), Carlos Levinho (Mdl)

june 2008

DOBES – file management

7

2 Brazilian National Prog.: ProDocLin

- Administration involves UNESCO
- Now 13 projects are selected and approved
- Digital equipment kits have been supplied, diffusing modern technology
- A first training course has been carried out in July 2009, upgrading national documentation capacity
- Instructors were principally:
 - Bruna Franchetto & Mara Santos (DOBES-Kuikuro)
 - Sebastian Drude (DOBES-Aweti)
 - Rosileide Costa (Museu Goeldi)

june 2008

DOBES – file management

8

3 LAT archives in Rio and Belém

- In 2007, an LAT server has been installed at the **Museu do Índio/FUNAI** in Rio de Janeiro (hence the ProDocLin national program)
- Another server has been set up in 2008/2009 at the **Museu Goeldi/MCT** in Belém do Pará
- Technical assistance from MPI Nijmegen was and is very useful — Thanks!
- Much legacy data has been and is being digitalized and is to be included in the Brazilian archives

june 2008

DOBES – file management

9

3 LAT archives in Rio and Belém

- The archive at Museu Goeldi in Belém has now the support of two larger projects, both coordinated by Vilacy Galúcio:
 - “Young Scientist” Project / CNPq (1 year) – acquisition of software & backup equipment, field trips and payment of services;
 - “Diffuse Rights Funds” Project / MCT (2 years) – basically for construction and maintenance of the archive by temporal employment of technicians
- Major problem: lack of permanent personnel

june 2008

DOBES – file management

10

3 LAT archives in Rio and Belém

Current situation of the archive in Belém:

- ~800 GB raw (uncut) data in HD-Storages
- ~500 CDs/DVDs still to be transferred
- many more A&V tapes are to be digitalized
- all data is still being classified and the formats standardized
- existing material exceeds by far the hardware capacity of
 - the HD-storages (2TB due to redundancy) &
 - the LAT server (1,8TB, with redundancy)

june 2008

DOBES – file management

11

3 LAT archives in Rio and Belém

The screenshot shows a web-based metadata management interface (WMI) for the ISLE Metadata Initiative. The interface is divided into several sections:

- Session:** Nome APP 20080713-CPA-museu, Título Museu da Aldeia Caribenha, Data 2009-07-17
- Description:** Projeto de Museu na Aldeia Caribenha do Povo Indígena Aikara do Pará
- Location:** Confed South-America, Country Brazil, Address Aldeia Caribenha - Para
- Project:** Projeto do CPDQ, Nome Projeto do CPDQ, Tipo Implementar o Centro de Documentação Permanente de Línguas e Culturas Indígenas da Amazônia no Museu Goeldi, M CPDQ
- Contact:** Ana Vilacy Galucio
- Description (Detailed):** O projeto busca promover uma documentação relacionada à preservação e arquivagem do patrimônio cultural brasileiro, especificamente através das línguas e culturas indígenas e de outros saberes da linha temática "Patrimônio Cultural Brasileiro". Os objetivos principais do projeto são: (i) implementar o Centro de Documentação de Línguas e Culturas Indígenas no Goeldi, no Museu Brasil Central, visando preservar o acervo linguístico do Círculo Científico de Humo Goeldi; (ii) desenvolver as línguas e culturas indígenas no Goeldi para fortalecer a memória, para arquivar e disponibilizar o patrimônio linguístico e cultural nacional; e (iii)
- Contact:** Centro Caribenha, Indígenas Caribenhos, RACK Caribenhos Indígenas, Museu do Índio, Subject, Interação interativa, Planning/Type operacional, Identificação não-útil, Social/Cultural (Interacted), Event/Structure Unpublished, Channel Page by Page

4 The PALIM - Portal (Wiki)

S. Drude is preparing "PALIM", a site with information on most technical aspects of language documentation:

- data types, formats, and standards,
- organization, administration, conventions,
- procedures, conversions, automatization,
- tools, manuals, infos, links, how-tos...

Currently: <http://palim.wikidot.com>

June 2008

DOBES - file management

13

4 The PALIM - Portal (Wiki)

Several similar / related sites exist, e.g.:

- [www.mpi.nl/DOBES/help .../documents](http://www.mpi.nl/DOBES/help.../documents)
- emeld.org/school/classroom/howto.html
- www.hrelp.org/archive/resources
- www.lat-mpi.eu/papers/papers-2005
- comp.ling.utexas.edu/wiki/doku.php/cindl

PALIM is in Portuguese and will have a focus on practical / technological matters, standards, how-tos

June 2008

DOBES - file management

14

4 The PALIM - Portal (Wiki)

I am interested in:

- Sites with information
- How-tos, manuals, specific settings...
- Solutions and setups that may be taken over by new documentation projects
- Requests: what should be there?

Later: a translation to English is desirable

June 2008

DOBES - file management

15

5 Brazilian national language survey

- Brazilian congress asked National Institute of Historical and Artistic Patrimony (IPHAN) to treat languages as immaterial patrimony
- the IPHAN workgroup (in spite of opposition) decided on a field survey of all languages of Brazil
- planned content of survey: —language name(s), —number of speakers, —degree of transmission, —recorded sample, etc.
- Hopefully a model for other countries

June 2008

DOBES - file management

16

5 Brazilian national language survey

Pilot projects to determine field methodology

- **1st. problem: Standardized field procedures**, for example:
 - Locate speakers, —Replicable results,
 - Computerized data sets (for data base)
- **2nd. problem: Costs**
 - 'Progressive' pilot: R\$10.000, 7 languages, 450 natives, 1 field person, 5 weeks, includes speakers off reserves
 - 'Old Guard' pilot: R\$228.000, 1 language, 300 natives in a reserve, 25 field persons, 1 year (would result in a cost of >R\$400.000.000 if done on a national scale)

June 2008

DOBES - file management

17

6 Languages in 2010 Brazilian census

- IBGE is preparing a new census for 2010
- One question about the languages spoken is to be included, only for indigenous respondents
- Census difficulties were resolved by the same IPHAN workgroup:
 - Names of ethnic groups will be used to identify languages, "I speak the language of the ___ group", since most native languages have no separate names.
 - In this way, the list of ethnic groups (with all variant names), already on the census taker's PDA, can identify the language, using a table of groups and languages.

June 2008

DOBES - file management

18

6 Languages in 2010 Brazilian census

The necessary data base adopted was that by D. Moore & Franchetto, avoiding the confusion between:

- languages vs. dialects
- population vs. number of speakers
- languages vs. ethnic groups.

It is essential to have correct information, since the database is official for federal programs (education, FUNAI use, etc.) and since a preliminary estimate of the degree of endangerment, which was greatly underestimated in the past, is needed for practical planning.

june 2008

DOBES – file management

19

7 Identification and classification

- The UNESCO Atlas of Endangered Languages (online) now contains data on Brazil
- Authors: D. Moore and B. Franchetto
- It is an updated and revised version of the data base for the IBGE census
- The Atlas makes a more accurate picture of the language situation widely available

june 2008

DOBES – file management

20

7 Identification and classification

Three major lists / classifications (there are others):

- Aryon Rodrigues: **171** (2006) to **205** (2008) indig. languages (lists each time more dialects as languages); population numbers are given as speaker numbers, drastically underestimating language endangerment
- Ethnologue / ISO: **226** indig. languages (not clear the definition of 'language'; in practice SIL/Ethnologue often lists separate dialects) of these 22 "extinct" and 33 with "no known speakers" -> 171 living lgs.
- Denny Moore et al: **154** languages plus ~36 major varieties, 12 of these 190 are hold to be extinct; tries to remedy confusion with varieties and ethnonyms, better speaker number data

june 2008

DOBES – file management

22

7 Identification and classification

- Language vs. dialect is mostly a question of definition, less often of lack of data
- Classification into families is generally clear
- Several language isolates or unclassified languages, and various uncontacted groups
- Different traditions for South America:
 - Greenberg and followers: All is ultimately AMERIND. Often exists strong counter-evidence against specific Amerind-internal macro groupings
 - Loukotka, Rodrigues, Kauffmann, Fabre, Moore, SIL etc.: conservative families; macro-groupings are considered speculative without more data

june 2008

DOBES – file management

23

7 Identification and classification

- With the national programs and the results of language documentation,
- and with many younger and well trained linguists emerging in Brazil,
- we are now or soon in a position to answer many questions and clear many doubts.
- We hold: the 639-3 ISO codes for identifying the languages should be in the hand of **academic** linguists, for example the same linguists who prepared the UNESCO Atlas.

june 2008

DOBES – file management

24