



**HAL**  
open science

## La Collection Pangloss : s'ouvrir à la science... et au reste du monde

Séverine Guillaume, Alexis Michaud, Balthazar Do Nascimento

► **To cite this version:**

Séverine Guillaume, Alexis Michaud, Balthazar Do Nascimento. La Collection Pangloss : s'ouvrir à la science... et au reste du monde. Symposium 2019 du Laboratoire d'excellence "Fondements empiriques de la linguistique / Empirical Foundations of Linguistics" (Labex EFL), Jun 2019, Paris, France. 2019. halshs-02156809v2

**HAL Id: halshs-02156809**

**<https://shs.hal.science/halshs-02156809v2>**

Submitted on 30 May 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License



# La collection Pangloss : S'ouvrir à la science... et au reste du monde

Séverine Guillaume\* Alexis Michaud\* Balthazar Do Nascimento\*

Contes, mythes, chansons, listes de mots... La collection Pangloss donne accès à des données de langues du monde entier, sous forme d'enregistrements audio ou vidéo, transcrits et traduits.

## LE PROJET

« Documentation fondamentale »

pour la sauvegarde, diffusion et valorisation de langues rares, souvent en danger

**Pérennisation** des documents

uniques, fruits d'explorations non reproductibles

Accueil des données de langues de toutes les aires géographiques et culturelles

(environ 20 langues en 2001, 70 langues en 2011, 191 en juin 2022)

The screenshot shows the Pangloss website interface. At the top, it says 'Qui sommes-nous? Corpus Dictionnaires'. Below that, there's a section for 'El Mejdub' with a story snippet: 'This story builds on the local legend that Abderrahman Lmejdub (El Majdoub) left the Senhaja people communicating with a mixed (or mixed-up) language. The story explains why this language is mixed with pieces of many languages.' There are 'Options d'affichage' (display options) for text and words. Below is a world map with numerous location markers indicating the geographical spread of the languages in the collection.

This block compares two types of data capture. On the left, a 'Manuscrit scanné' (scanned manuscript) shows handwritten text in a non-Latin script. On the right, a 'Vidéo annotée' (annotated video) shows a person speaking, with a subtitle overlay: 'What's the name, the snake there, ndrî? ?mî'ke' sa' ?u'je' la?'. This illustrates the transition from physical records to digital, annotated audio-visual data.

This block shows a screenshot of the 'Lexica' website, which lists dictionaries from the Pangloss collection. The table includes columns for 'Titre' (Title), 'Langue' (Language), and 'Chercheur(s)' (Researcher(s)).

Titre	Langue	Chercheur(s)
Dictionnaire Japhug	Japhug	Guillaume Jacques (CNRS)
Dictionnaire Khasang	Khasang	Guillaume Jacques (CNRS) et Anna Giacalone (IFILC)
Dictionnaire Limbu	Limbu	Boyd Michalovec (LACITO)
Dictionnaire Meeleq	Meeleq	Alexandre François (LACITO)
Dictionnaire Na	Na	Alexis Michaud (LACITO)
Dictionnaire Yungva	Yungva	Séverine G. (LACITO)

## À QUOI ÇA SERT?

→ à chacun d'en décider !

- enseignement : perpétuer les langues et cultures
- libre usage par les communautés de locuteurs
- en Traitement Automatique des Langues : défis techniques (transcription automatique, ...)

*liste ouverte !*

This block contains the logos of the project partners: Pangloss (pangloss.cnrs.fr), CNRS, Huma-Num, LACITO (Langues et Civilisations à Tradition Orale), and COCOon. A QR code is also present for more information.

\*laboratoire Langues et Civilisations à Tradition Orale (LACITO), Unité Mixte de Recherche CNRS / Sorbonne Nouvelle / INALCO

