



**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, there's a navigation bar with 'Qui sommes-nous?', 'Corpus', and 'Dictionnaires'. The main content area features a story titled 'El Mejdub' with a small video player. Below the story, there are 'Options d'affichage' (display options) for text and words, with checkboxes for 'autre', 'en', and 'autre'. A world map at the bottom shows various location markers across different continents, indicating the global reach of the collection.

This block compares two types of data collection. On the left, a 'Manuscrit scanné' (scanned manuscript) shows handwritten text in a non-Latin script on a grid background. On the right, a 'Vidéo annotée' (annotated video) shows two people in a conversation with a subtitle in a non-Latin script at the bottom.

The screenshot shows the 'Lexica' section of the Pangloss website, titled 'les dictionnaires de la collection Pangloss'. It lists several dictionaries with columns for 'Titre', 'Langue', and 'Chercheur(s)'. The list includes dictionaries for Japhug, Khasang, Limbu, Meebing, Na, and Yungba, each with a small icon and the name of the researcher(s).

## 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', 'COCOon', and 'LACITO (Langues et Civilisations à Tradition Orale)'. A QR code is located on the right side of the block.

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

