

**Collection Pangloss :**

Langues principalement à tradition orale  
**170 langues, 745 heures d'enregistrements**  
 (720h d'audio et 25h de vidéo),  
**1700 documents (230h) transcrits, traduits, glosés**



Documents recueillis par des linguistes de terrain «généralistes» dans le monde entier



- **Documentation fondamentale** pour la sauvegarde, diffusion et valorisation de langues minoritaires, en danger
- **Pérennisation** de documents uniques, fruits d'explorations non reproductibles (archivage pérenne au CINES via les services de la TGIR Huma-Num)
- Accueille des données de **langues de toutes les aires géographiques et culturelles** (environ 20 langues en 2001, 70 langues en 2011, 170 à ce jour)

#### Extrait XML des annotations de l'enregistrement audio

```
<TEXT xml:lang="lhu" id="crdo-LHU_LAHU_JM_20">
<HEADER>
<TITLE xml:lang="en">How We Came from Burma</TITLE>
<SOUNDFILE href="Thailand/Lahu/LAHU_JM_20.wav"/>
</HEADER>
(...)
<S id="LAHU_JM_20s3">
<AUDIO start="15.9382" end="20.0724"/>
<FORM kindOf="phono">yâ kâ? o-mô=lô mâ ve... tô ve...</FORM>
<TRANSL xml:lang="en">And there was a big flock of kids along too--er, walking along--.</TRANSL>
<W>
<FORM kindOf="phono">yâ</FORM>
<TRANSL xml:lang="en">children</TRANSL>
<M>
<FORM kindOf="phono">kâ?</FORM>
<TRANSL xml:lang="en">children</TRANSL>
</M>
</W>
<W>
<FORM kindOf="phono">kâ?</FORM>
<TRANSL xml:lang="en">PRT(also)</TRANSL>
<M>
<FORM kindOf="phono">kâ?</FORM>
<TRANSL xml:lang="en">PRT(also)</TRANSL>
</M>
</W>
<W>
<FORM kindOf="phono">o-mô=lô</FORM>
<TRANSL xml:lang="en">big.group</TRANSL>
<M>
<FORM kindOf="phono">o</FORM>
<TRANSL xml:lang="en">PREF</TRANSL>
</M>
<M>
<FORM kindOf="phono">mô</FORM>
<TRANSL xml:lang="en">group</TRANSL>
</M>
<M>
<FORM kindOf="phono">lô</FORM>
<TRANSL xml:lang="en">big</TRANSL>
</M>
</W>
(...)</S>
</TEXT>
```

#### Interface de consultation en ligne

S1 [yâ-hi Man-mi-gi chê ve, chô Thây=mô-mi 5 pho la ve yâ?cp há-qá pho la ve tê yâ, phô? tu ca cå, nà tå cå là ve, chô 5 gâ la ve.]  
 1p.-pluralizer Burma-earth-skin stay PRT(NOM) here Thailand-heaven-earth PRT(LOC) flee PRT(CIS) PRT(NOM) road side flee PRT(CIS) PRT(NOM) one time young plant  
 ca cå nå tå cå là ve chô 5 gâ la ve  
 look.for eat fish dam eat come PRT(NOM) here PRT(LOC) arrive PRT(CIS) PRT(NOM)  
 [When we were] fleeing here to Thailand [from where] we [used to] live in Burma, when we were fleeing here along the roads and byways, we lived off jungle-greens and by damming [streams] for fish as we came—that's how we made it here.

S2 [yâ kâ? 3-mô-lô mâ ve... tô ve...]  
 children PRT(also) PREF-group-big numerous PRT(NOM) walk PRT(NOM)  
 And there was a big flock of kids along too--er, walking along--.

S3 [yâ-hi yâ?cp-qâ kâ? gâ chê mo jâ.]  
 1p.-pluralizer road-side PRT(also) have.to stay.be.a.long.time very  
 And we had to stay on the road a very long time too.

#### Technologies de la parole pour langues peu dotées : un domaine de recherche très actif

- Enrichir la base de données au moyen d'outils numériques
  - Alignement temporel texte-son par alignement forcé
  - Etiqueteurs morphosyntaxiques pour génération automatique de gloses
- Défi scientifique du scénario « ressources limitées » pour la recherche en informatique
  - Transcription automatique de la parole  
Initiatives récentes : wav2letter++, persephone, elpis...
  - Synthèse de la parole
  - Traduction automatique
- Potentiel pour l'enseignement : outils ludiques grand public

