La collection Pangloss : S'ouvrir aux langues rares. . . et aux autres cultures du monde

Balthazar Do Nascimento*¹, Séverine Guillaume¹, and Léa Mouton¹

¹UMR 7107 LACITO – Centre national de la recherche scientifique - CNRS (France), University Paris Sorbonne, INALCO – France

Résumé

La collection Pangloss est une archive de langues rares (plus de 200 langues), constituée d'enregistrements sonores ou de vidéos accompagnés d'annotations (transcription, traduction, gloses). Toutes les ressources (> 5500) sont sous licence Creative Commons et sont librement accessibles en ligne.

Ainsi, légendes, récits de vie ou d'artisans, listes de vocabulaire... sont disponibles pour les usages les plus divers : recherche, enseignement, ou simple découverte.

Les utilisateurs de ces données sont également invité.e.s à contribuer à l'enrichissement de l'annotation de ces corpus, et à explorer de nouvelles directions pour leur exploitation. Les corpus de la collection Pangloss donnent ainsi lieu à des projets de recherche comme par exemple, la transcription automatique de la parole.

Mots-Clés: Pérennisation, Documentation fondamentale, Corpus oraux, Traitement automatique des langues, Enseignement, XML, annotation, Archive ouverte, Donnée linguistique de terrain

*Intervenant