



COMMUNIQUÉ DE PRESSE NATIONAL - PARIS - 14 JANVIER 2021

Patrimoine linguistique en danger : un nouvel écrin pour la collection Pangloss

Comme certaines espèces animales et végétales, une partie des langues du monde est en danger d'extinction. Initiée en 1995 par le laboratoire Langues et civilisations à tradition orale (CNRS/ Université Sorbonne Nouvelle/Inalco), la collection Pangloss réunit des enregistrements de ces langues en danger. En effet, celles qui n'ont pas de tradition écrite (l'immense majorité) disparaissent complètement avec leurs derniers locuteurs. Cet effort de sauvegarde et de mise à disposition du patrimoine linguistique s'étend aussi à d'autres langues rares, relativement peu documentées. Grâce au soutien du CNRS, la collection Pangloss fait aujourd'hui peau neuve avec un nouveau site web qui s'adresse aussi au grand public.

À ce jour, la bibliothèque sonore Pangloss contient plus de 3600 enregistrements audio ou vidéo en 170 langues de tous les continents. On y trouve ainsi des contes et chansons en xârâgurè (Nouvelle-Calédonie), des conversations et des récits en kakabe (Guinée), des recettes de cuisine en koyi rai (Népal) et en na-našu (Italie)... soit 780 heures d'écoute au total.

Ces extraits sonores sont le fruit de plus de vingt ans de travail de linguistes et d'ethnologues qui, chacun sur leur terrain d'étude, œuvrent à la collecte et à la sauvegarde du patrimoine linguistique mondial. Certains sons proviennent de la numérisation d'anciennes bandes magnétiques¹. Près de la moitié des enregistrements sont transcrits et annotés, par exemple avec des éléments de contexte ou des traductions vers d'autres langues. Le site est ouvert aux contributions d'experts, académiques ou non, pour améliorer le corpus en participant aux transcriptions et traductions.

Avec son design repensé, le site pangloss.cnrs.fr peut désormais être consulté avec deux niveaux de lecture, afin d'être plus accessible au grand public qui peut librement écouter et télécharger ces témoignages de la diversité linguistique. En grande partie sous licence *Creative Commons*, les contenus pourraient notamment alimenter des projets muséographiques ou des créations sonores.

Au-delà de son aspect patrimonial, cette collection participe d'une démarche de science ouverte, en facilitant la conservation, le référencement² et la mise à disposition des données primaires des chercheurs et chercheuses. Elle entend ainsi lutter contre la déperdition des données scientifiques (une « deuxième mort » pour les langues disparues) mais aussi favoriser les collaborations avec d'autres disciplines : les informaticiens intéressés par le traitement automatique des langues peuvent y trouver facilement les fichiers dont ils ont besoin et participer au co-développement d'outils (de transcription automatique, par exemple). Entièrement bilingue français-anglais, le site comprend des traductions partielles dans d'autres langues, dont le chinois pour les notices de certaines langues asiatiques.

Outre les contributions de différents laboratoires associés au CNRS³, la collection Pangloss est soutenue par l'Institut des langues rares de l'EPHE-PSL, récemment créé. Les données sont sauvegardées dans l'archive de la Très grande infrastructure de recherche (TGIR) Huma-Num. La collection Pangloss est membre du réseau international DELAMAN, Digital Endangered Languages and Musics Archives Network. Elle est hébergée par la plateforme Cocoon, Collection de corpus oraux numériques, qui participe au réseau international OLAC, Open Language Archive Community.



Notes

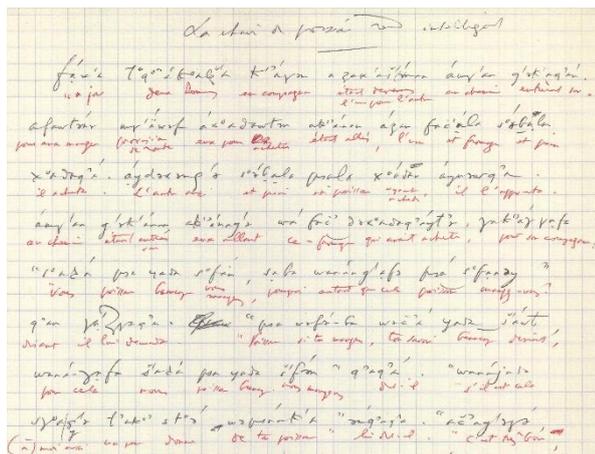
¹ Comme ceux du dernier locuteur de la langue oubykh, par Georges Dumézil, en 1968.

² Chaque ressource est dotée d'identifiants de types DOI (*Digital Object Identifier*) et ARK (*Archival Resource Key*).

³ En particulier (liste non limitative) : Langues et civilisations à tradition orale (Lacito, CNRS/Université Sorbonne Nouvelle/Inalco) ; Centre de recherches linguistiques sur l'Asie orientale (CRLAO, CNRS/Inalco/EHESS) ; Langage, langues et cultures d'Afrique noire (Llacan, CNRS/Inalco) ; Structure et dynamique des langues (Sedyl, CNRS/Inalco/IRD).

Quelques exemples de contenus à découvrir :

- « La chair de poisson rend intelligent », une histoire en oubykh (langue caucasienne autrefois parlée en Turquie et Géorgie, qui compte quelque 80 consonnes) racontée par Tevfik Esenç, son dernier locuteur, et enregistrée par le linguiste et anthropologue Georges Dumézil en 1968 (ses notes manuscrites sont aussi consultables). Lien : <https://doi.org/10.24397/pangloss-0004320>
- Corpus audio et vidéo en kakabe, langue de Guinée (qui fut par le passé une langue d'esclaves ou de serviteurs), enregistré et déposé par la linguiste Alexandra Vydrina. Lien : <https://pangloss.cnrs.fr/corpus/Kakabe>



Transcription manuscrite, par Georges Dumézil, d'un récit oubykh raconté par Tevfik Esenç. © Georges Dumézil



Les femmes du village Kouroupampa (Guinée) faisant la cuisine (photo issue du corpus en langue kakabe). © Alexandra Vydrina

Deux articles pour aller plus loin :

[Une archive ouverte pour sauvegarder le patrimoine linguistique mondial](#)

[Pangloss à l'écoute des langues rares](#)

Contacts

Chercheur CNRS | Alexis Michaud | alexis.michaud@cnrs.fr

Ingénieure CNRS | Séverine Guillaume | severine.guillaume@cnrs.fr

Presse CNRS | Véronique Etienne | T +33 1 44 96 51 37 | veronique.etienne@cnrs.fr