

Utilisation du logiciel ARBIL pour gérer les métadonnées d'enregistrements audio annotés

ARBIL est un logiciel développé par le Max Planck Institut de Nijmegen aux Pays-Bas. Il permet la gestion des métadonnées de différents types de documents, et particulièrement d'enregistrements audio ou vidéo issus d'enquêtes de terrain, ainsi que des textes d'annotation de ces ressources. Il est téléchargeable à cette adresse :

<http://tla.mpi.nl/tools/tla-tools/arbil/>

Windows : Download the Windows installer

Mac : Download the Mac installer

L'installation se fait simplement en double-cliquant sur le fichier téléchargé.

Lorsque vous lancez **ARBIL** la première fois, il vous sera demandé si vous souhaitez l'utiliser en *IMDI* ou *CMDI*. Actuellement nous utilisons le format *IMDI* (il existe un convertisseur *IMDI* > *CMDI*).

Choisir le dossier Local Corpus Storage Directory

Par défaut, ARBIL utilise pour enregistrer les fichiers de métadonnées, un dossier de travail situé dans votre dossier système *utilisateur*. Vous pouvez choisir un autre lieu (par exemple sur une partition *Data* (D:) si vous en avez une)

Options, Local Corpus Storage Directory

Move Local Corpus Storage Directory

Choisissez l'emplacement de ce dossier sur votre disque dur (dans le but de faciliter une sauvegarde régulière de vos données sur un support externe).

L'écran de ARBIL présente deux colonnes. Celle de droite est la zone de saisie des métadonnées, celle de gauche comporte trois sections : *Remote corpus*, *Local Corpus* et une section *outillage* en bas.

- ❖ **La section *Remote corpus*** vous permet **si vous avez une liaison Internet** de voir les ressources accessibles dans différents corpus gérés par le Max Planck Institute.

Par exemple en double-cliquant sur :

IMDI-corpora, puis Dobes Archive, Bainouk, Bainouk Gubeeher, Plants, Interviews

vous trouverez une session nommée : *DJI121109AC3* qui concerne deux 'consultants' (Actors : *Biagui et Sagna*) qui parlent de *l'histoire et la culture Bainouk* (content), dans le cadre d'un projet *Dobes* dirigé par *Frederik Lupke*. cette session a donné lieu à un enregistrement audio *DJI121109AC3.wav* de 42 Mo.

- ❖ **La section *Local corpus*** vous permet de gérer en local, sur votre machine, vos propres métadonnées.

Pour cela vous pouvez créer plusieurs niveaux hiérarchiques pour les ranger. Le niveau supérieur est de type *Corpus*. Il peut contenir d'autres corpus ou bien des sessions. Une *Session* correspond à une séance d'enregistrement, c'est-à-dire un lieu, une date, des intervenants et des enregistrements audio ou/et vidéo. Généralement on associe à une session, lorsque le travail aura été fait, les fichiers d'annotations (transcription, traduction, analyse morpho-syntaxique) ayant pour source les enregistrements audio/vidéo. Les métadonnées sont donc attachées à une session.

Remarque: Si vous avez des enregistrements audio de longue durée, vous pouvez les découper en plusieurs fichiers audio qui constitueront autant de (sous-)sessions.

Il est conseillé de choisir un *modèle de nommage* de vos fichiers audio/vidéo dans le cadre d'un projet. Par exemple dans le projet *Sénélangues*, le modèle de nommage est :

codeLangue_IntialesChercheurs_TypeEnregistrement_numéroEnregistrement_motClé

ainsi : *CAE_ED_NARR_04_MOON.WAV* et *CAE_ED_NARR_04_MOON.EAF*

représenteront respectivement le fichier audio (WAV) et le fichier d'annotation ELAN (EAF) d'un enregistrement (le 4ème dans la série : *04*) de type narration (NARR) en langue laalaa (*CAE*) de El Hadji D. (*ED*) et qui parle de la lune (*MOON*).

Créer un nouveau Corpus (par exemple Sénélangues)

Clic-droit sur Local Corpus, **Add, Corpus**

Un *Standard corpus* a été créé. Cliquez dedans, et dans la fenêtre d'édition de droite

Remplacez : *Standard Corpus* par *Sénélangues*

Dans ce corpus, vous pouvez maintenant créer des *sessions* dans lesquelles vous allez décrire des sessions d'enregistrement audio (ou vidéo).

Créer une nouvelle Session dans un corpus, faites dans la colonne de gauche:

Clic-droit sur le nom du corpus (*Sénélangues*), **Add, Session**

Une session *Standard (Session) Resource Bundle* est créée.

Saisir les métadonnées de la session

Pour accéder aux métadonnées de la Session, double-cliquez dessus. Les noeuds *Content* et *Project* apparaissent sous la Session et, dans la fenêtre de droite, une table de métadonnées relatives à la Session s'affiche.

Renommez la Session *Standard (Session) Resource Bundle* créée en lui donnant (c'est un conseil) le même nom que celui du fichier audio qu'elle est sensée décrire (**sans** l'extension WAV). Saisissez un *Titre* pour cette session, la *Date*, éventuellement une description de la session, ainsi que le lieu (*Continent, Country, Region*) où elle s'est tenue.

Saisir les métadonnées des noeuds Content et Project de la session

Pour accéder aux métadonnées du noeud *Content* ou *Projet*, double-cliquez dessus. Dans la fenêtre de droite les métadonnées relatives à ce noeud apparaissent et peuvent être remplies. Les lignes présentant une icône [CV] à sa droite peut être double-cliquée pour faire apparaître une liste déroulante permettant de choisir une valeur prédéfinie. Si l'icône [CV] est encerclée d'un trait discontinu, on peut saisir une valeur autre que celles prédéterminées.

Importer dans votre corpus un fichier de métadonnées au format IMDI

Vous pouvez importer dans votre corpus des fichiers de métadonnées au format IMDI

par exemple, copier le fichier ci-joint *SAY_BC_INT_01_Religion.imdi* sur votre machine. Puis faites :

Clic-droit sur le nom de votre corpus (*Sénélangues*) dans la colonne de gauche,

Import Branch, rechercher le fichier *SAY_BC_INT_01_Religion.imdi* puis faites **Start**, et enfin **Close**

La session *SAY_BC_INT_01_Religion.imdi* s'installe dans le corpus *Sénélangues*.

Vous pouvez vous servir d'une session importée comme *modèle* pour saisir les métadonnées d'une de vos sessions, en remplaçant les données actuelles par les vôtres.

Pour cela, après avoir importé la session modèle, commencez par cliquer sur le nom de la session nouvellement importée et corrigez tout ce qui doit l'être dans la zone d'édition de droite. Puis vous ouvrez un à un les noeuds (en double-cliquant sur chaque noeud la colonne de gauche) et vous les corrigez.

Faites attention de bien visiter et corriger toutes les noeuds-enfants des 4 noeuds principaux : **Actors**, **Content**, **MediaFiles**, **WrittenResources**. (Le noeud **Project** ne devrait pas être modifiée, il concerne le Projet *Sénélangues* dans son ensemble.)

Ajouter des noeuds à une session

Les métadonnées relatives à une session d'enregistrement de données de terrain comporte généralement les noeuds suivants :

- **Projet** : décrivent le projet qui a occasionné cette session (ce peut-être un projet personnel)
- **Content** : décrivent le genre, les modalités, le sujet... de la session
- **Actors** : informent sur les participants à la session (chercheur, informateurs...), leur âge, statut social... les langues qu'ils parlent
- **Le(s) fichier(s) audio et/ou vidéo enregistré(s) à l'occasion de cette session**
- **les fichiers d'annotation (transcription, traduction, analyse morpho-syntaxique...) qui en sont issus**

Pour ajouter des noeuds à une session, par exemple pour un locuteur (**Actor**)

Clic-droit sur la icône de la **Session**, puis **Add Actor**

De même, chaque noeud peut comporter différents noeuds enfants en procédant de la même manière, par exemple un autre participant (*Actor*) à la Session,

Clic-droit sur le noeud **Actors**, **Add, Actor**

Idem pour les différentes langues parlées par chaque participant actif (clic-droit sur l'**Actor**, **Add, Language**). *Remarque*: Les noeuds que l'on peut ajouter dépendent du noeud parent (par exemple, on ne peut pas ajouter un noeud *Actor* dans la noeud *Projet*).

Supprimer un noeud

Un noeud (et ses noeuds enfants) peut-être supprimé simplement :

Clic-droit sur le noeud, **delete**

❖ La sous-section *Files*

La troisième zone de la colonne de gauche de Arbil comporte 2 onglets : *Files* et *Favourites*.

Files contient un dossier *Working directories* dans lequel vous pouvez ajouter les dossiers de votre machine qui contiennent les fichiers audio et vidéo à associer aux différentes Sessions.

Ajouter des dossiers à *Working Directories*

Clic-droit sur *Working directories*, **Add Working directory**

Recherchez le dossier à lier

Cliquez sur *Add Working Directory*

En double-cliquant sur le dossier lié, la liste de ses fichiers apparaît. Une icône apparaît devant chaque fichier indiquant sa nature (audio, vidéo, autre) et s'il est conforme au standard.

Ajouter un *MediaFile* ou un *WrittenResource* à une session par glissé-déposé

Une session se doit de contenir un fichier audio ou vidéo (*MediaFile*) et contiendra généralement un fichier d'annotation (*WrittenResource*) par exemple ELAN ou PDF...

Glissez-déposez le fichier voulu depuis la zone *Working directories* vers la Session

Un noeud *MediaFile* ou *WrittenResource* est créé sous la session. En cliquant sur le noeud, une table de métadonnées s'affiche pré-remplie avec une partie des informations relatives au fichier : nom, taille, format...

❖ La sous-section *Favourites*

Une fois que vous avez enregistré (Ctrl/s) votre session, vous pouvez ouvrir la zone tampon appelé *Favourites* en bas dans la colonne de gauche, en cliquant sur l'onglet *Favourites*. Cette zone sert à conserver des copies d'éléments de métadonnées afin de pouvoir les ré-utiliser ailleurs.

Par exemple, vous pouvez glissez/déposez la session entière que vous venez de créer. Vous pourrez ensuite vous servir de cette copie pour créer une nouvelle session dans votre corpus. Pour cela :

Glisser/déposer la Session de la zone *Favourites* sur le nom du corpus *Sénélangues* au dessus.

Ensuite vous modifierez les métadonnées de cette copie pour qu'elles correspondent à votre nouvelle session.

Vous pouvez aussi ne transférer dans *Favourites* que les métadonnées d'un locuteur (s'il revient régulièrement dans toutes vos sessions par exemple). Ou encore y transférer les métadonnées *Projet* pour pouvoir les ré-utiliser ensuite dans une nouvelle session sans avoir à re-saisir les informations sur le projet. *Remarque*: Dans ce dernier cas, il faudra auparavant supprimer (clic-droit, delete) le *Projet* vierge créé automatiquement lors de la création d'une nouvelle session, car une session ne peut contenir 2 projets.

N'oubliez d'enregistrer avec Ctrl/s lorsque vous avez fini.

Transmettre vos métadonnées à un correspondant

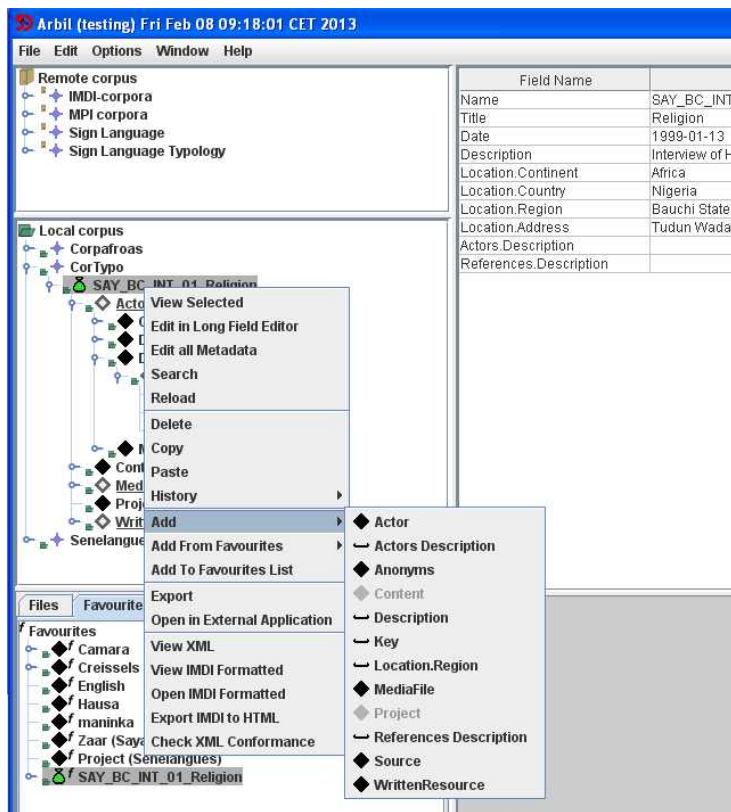
Clic-droit sur le nom de votre corpus (Sénélangues), Export

Choisissez un dossier où exporter ce corpus, puis faites **Start**.

Transmettez les fichiers (imdi) qui se trouvent maintenant dans le dossier choisi pour l'exportation. Si vous avez exporter plusieurs fois le même fichier au même endroit, des versions numérotées sont créées. Transmettez la dernière version.

Exemple de l'ajout d'un informateur à une Session

Clic-droit sur la Session, Add, Actor



Exemple d'ajout d'une langue à un informateur

Clic-droit sur l'informateur Dariya, Add, Actor Language

