

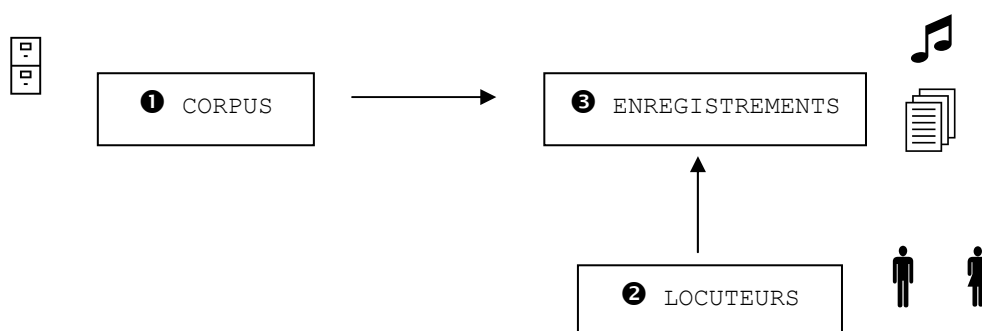
MOCA – Multimedia Oral Corpora Administration

Un système de gestion et d'annotation de données orales

Peter Gilles (Université du Luxembourg) peter.gilles@web.de
Laurent Kevers (CENTAL / UCLouvain) kevers@tedm.ucl.ac.be
Anne Catherine Simon (VALIBEL / UCLouvain) simon@rom.ucl.ac.be

<http://cental.fltr.ucl.ac.be:9080/moca>

1. Structure de la base de données... et historique



2. Fonctions pour l'utilisateur

[SOCIO] [TRANSCRIPT] [LABELS] [ADD DATA] [PREFERENCES] [LISTS]

- [socio] sélectionner des enregistrements selon des critères sur les locuteurs ou sur les enregistrements
- créer des sous-corpus [préférences] et les partager [administration]
- [transcript] lire des transcriptions alignées; rechercher dans les transcriptions
- [labels] annoter des transcriptions et rechercher dans les annotations
- [add data] encoder de nouvelles données sur les corpus, les locuteurs ou les enregistrements
- archiver des "objets" liés aux enregistrements [fiche "enregistrement"]

3. Administration

Les fonctions d'administrateur concernent l'installation et l'administration de la base de données et la gestion des utilisateurs.

Ajout dans une table annexe	Modifier une table annexe	Gestion utilisateurs	Upload transcript	Attribuer des enregistrements	Attribuer un vcorpus	Supprimer une fiche
-----------------------------	---------------------------	----------------------	-------------------	-------------------------------	----------------------	---------------------

- les tables annexes: définition; actions [ajout, modifier une table annexe]
- administration des utilisateurs
 - o profils d'utilisateurs: définitions; restrictions sur les actions
 - o restrictions sur les données [subset]
 - o créer un utilisateur (modifier son profil, son mot de passe) [gestion des utilisateurs]
 - o attribuer un sous-ensemble d'enregistrements [attribuer des enregistrements]
 - o attribuer des v-corpus [attribuer un v-corpus]
- chargement des sons et des transcriptions [upload transcript]
 - o copier les fichiers sons sur le serveur (via SSH ou FTP) (dénomination du fichier)
 - o créer une transcription provisoire (pour annoter du son non transcrit)
 - o uploader une transcription existante
 - le textgrid (format standard; nom du fichier et des locuteurs)
 - la transformation du textgrid
 - l'upload d'une transcription
 - le problème des corrections ultérieures

4. Installation de MOCA et personnalisation

- environnement nécessaire: serveur, système d'exploitation, logiciels (Apache, serveur web; Mysql, base de données; php pour les scripts et pages dynamiques; programmes annexes: Praat, scripts de transformation; logiciel d'accès aux répertoires du serveur: SSH, FTP...)
- créer un répertoire pour le site et y recopier tout le système avec les sous-répertoires et les fichiers
- création de la base de données (vide, sauf un utilisateur de type administrateur)
- personnaliser le fichier de configuration
 - o renseignements pour se connecter à la base de données
 - o chemins d'accès vers les répertoires (sons, transcriptions)
 - o chemins d'accès vers Praat, répertoires temporaires
 - o les constantes
 - valeurs des tables annexes (état civil, niveau d'études, localisations...)
 - fichier langues; nom de la base...
- remplir les tables annexes via l'interface web
 - o Certaines valeurs sont définies dans le fichier de constantes car elles représentent un ensemble fermé et sont dans une relation (1-1) avec une fiche
 - o Par contre: *langues, professions, localisation*: modifiables via l'interface, car ces catégories peuvent évoluer de manière périodique, et on peut attribuer des valeurs multiples à une fiche (1-N)